U.S. DEPARTMENT OF COMMERCE
National Technical Information Service

PB-272 274

Development of a System for Conducting Inter-Laboratory Tests for Water Quality and Effluent Measurements

FMC Corp, San Jose, Calif Advanced Products Div

Prepared for

Environmental Monitoring and Support Lab -Cincinnati, Ohio

Jun 77

PB 272 274

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U. S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA-600/4-77-031	2.	3. RECIPIENT'S ACCESSION NO. PB272274
4. TITLE AND SUBTITLE Development of a System for Conducting Inter-Laboratory Tests for Water Quality and Effluent Measurements		5. REPORT DATE June 1977 Issuing Date 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Arthur C. Green Robert Naegele		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS FMC Corporation 1105 Coleman Ave. San Jose, CA 95108		10. PROGRAM ELEMENT NO. 24AUB TASK NO. 5 11. CONTRACT/GRANT NO. 68-03-2115
12. SPONSORING AGENCY NAME AND ADDRESS Environmental Monitoring & Support Laboratory-Cin.,OH Office of Research and Development U.S. Environmental Protection Agency Cincinnati, Ohio 45268		13. TYPE OF REPORT AND PERIOD COVERED July 16, 1974 - April 15, 1976 14. SPONSORING AGENCY CODE EPA/600/06
15. SUPPLEMENTARY NOTES		

16. ABSTRACT

FMC Corporation has Developed a system for evaluating water pollution data and the laboratories which produce these data. The system consists of a plan for the design and implementation of an interlaboratory test program. A pilot test program was included to evaluate and to verify the complete program.

Investigation of ongoing interlaboratory testing programs were conducted and their deficiencies identified in their design and in the procedures by which they were conducted. The conclusions and recommendations presented in the report are support by an extensive literature review of previous interlaboratory tests and their methods for experimental design and test data analyses. Additionally, 18 EPA, State, and private laboratories were visited to review their comments regarding difficulties and deficiencies in interlaboratory test programs in general.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Laboratories Chemical Laboratories Quality Control Acceptable Quality Level Reproducibility Statistical Tests		07B
18. DISTRIBUTION STATEMENT Release to Public	19. SECURITY CLASS *(This Report)* Unclassified	21. NO. OF PAGES 144
	20. SECURITY CLASS *(This page)* Unclassified	22. PRICE PCA07 / MF A01

EPA Form 2220-1 (9-73)

☆U.S. GOVERNMENT PRINTING OFFICE: 1977—757-056/6473

EPA-600/4-77-031
June 1977

DEVELOPMENT OF A SYSTEM FOR CONDUCTING INTER-LABORATORY TESTS FOR WATER QUALITY AND EFFLUENT MEASUREMENTS

by

Arthur C. Green
Robert Naegele

FMC Corporation
Advanced Products Division
San Jose, California 95108

Contract 68-03-2115

Project Officer
Terry C. Covert
Environmental Monitoring and Support Laboratory
Cincinnati, Ohio 45268

ENVIRONMENTAL MONITORING AND SUPPORT LABORATORY
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
CINCINNATI, OHIO 45268

DISCLAIMER

This report has been reviewed by the Environmental Monitoring and Support Laboratory, U.S. Environmental Protection Agency and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the U.S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

FOREWORD

Environmental measurements are required to determine the quality of ambient waters and the character of waste effluents. The Environmental Monitoring and Support Laboratory - Cincinnati, conducts research to:

- Develop and evaluate techniques to measure the presence and concentration of physical, chemical, and radiological pollutants in water, wastewater, bottom sediments, and solid waste.

- Investigate methods for the concentration, recovery, and identification of viruses, bacteria and other microbiological organisms in water. Conduct studies to determine the responses of aquatic organisms to water quality.

- Conduct an Agency-wide quality assurance program to assure standardization and quality control systems for monitoring water and wastewater.

In carrying out its legislated mandates, U.S. EPA requires water quality and effluent monitoring data from a broad spectrum of laboratory and field operations--federal, state, local, contract and private. The quality (precision and accuracy) of the data generated by the various monitory activities must be known if EPA is to use these data to assess pollution trends, set standards, verify compliance with regulations, and conduct enforcement actions.

In order for EPA to significantly improve its capability to assess the validity of the data it receives and uses, its interlaboratory testing program must be substantially expanded. This report developed by FMC Corporation contains a formalized system to assure that all laboratories and all necessary measurements are continually evaluated as to their performance and reliability. The report contains a systematic plan for conducting interlaboratory tests for water pollution measurements and establishes their relationship to the overall external quality control evaluation program.

This report is not an official EPA manual. Rather, it is a research report that is but one of a series being used as an input to develop *EPA Manuals and Guidelines*.

DWIGHT G. BALLINGER, Director
Environmental Monitoring &
Support Laboratory/Cincinnati

ABSTRACT

FMC Corporation has developed a system for evaluating water pollution data and the laboratories which produce these data. The system consists of a plan for the design and implementation of an interlaboratory test program. A pilot test program was included to evaluate and to verify the complete program.

Investigation of ongoing interlaboratory testing programs were conducted and their deficiencies identified in their design and in the procedures by which they were conducted. The conclusions and recommendations presented in the report are supported by an extensive literature review of previous interlaboratory tests and their methods for experimental design and test data analyses. Additionally, 18 EPA, state and private laboratories were visited to receive their comments regarding difficulties and deficiencies in interlaboratory test programs in general.

This report was submitted in fulfillment of Contract No. 68-03-2115 by FMC Corporation under the sponsorship of the U.S. Environmental Protection Agency. This work covers a period from July 16, 1974, to April 15, 1976.

CONTENTS

Page

Foreword.... iii
Abstract.... iv
Figures.... vi
Tables.... vii
Abbreviations and Symbols.... viii
Acknowledgements.... x

I Introduction.... 1

II Summary.... 4

III Conclusions.... 5

IV Recommendations.... 7

V Literature Survey.... 8

VI Field Investigation.... 29

VII Data Analysis and Evaluation.... 34

VIII Interlaboratory Test - Program Plan.... 46

IX Pilot Program.... 56

X References.... 89

XI Interlaboratory Test Programs - Bibliography.... 98

XII Appendix.... 127

FIGURES

Number Page

5-1 Percent of Insoluble Residue.................................. 17
7-1 Information Flow Model.. 35
7-2 Error Diagram... 37
7-3 Sources of Error in Measurement............................... 38
7-4 Normality Test and Outliers Treatment......................... 43
8-1 Inter-Laboratory Test Program................................. 48
9-1 Youden's Plot, Cu, Samples 1 & 4.............................. 70
9-2 Youden's Plot, Cu, Samples 2 & 3.............................. 71
9-3 Youden's Plot, Cu, Samples 5 & 6.............................. 72
9-4 Youden's Plot, Zn, Samples 1 & 4.............................. 73
9-5 Youden's Plot, Zn, Samples 2 & 3.............................. 74
9-6 Youden's Plot, Zn, Samples 5 & 6.............................. 75
9-7 Relative Errors Distribution, Cu.............................. 77
9-8 Relative Errors Distribution, Zn.............................. 78
9-9 Thompson's Ranking Scores for 16 Laboratories................. 85
9-10 Mean Score of (M-n) Laboratories............................. 86

TABLES

Number		Page
5-1	Median Values of Sample Constituents (Table I of Reference 46)	19
5-2	Water-Insoluble Nitrogen Results (Table 6 of Reference 1)	20
5-3	Data for Two Difference Methods	24
5-4	Eight Combinations of Seven Factors Used to Test Ruggedness of an Analytical Method (Table 8 of Reference 1)	26
5-5	Measurement of H_2O in Phosphoric Acid	27
6-1	Agencies Visited During Field Investigation	32
7-1	Lab Training and Data Evaluation	45
9-1	Sample Compositions for Pilot Test Program	57
9-2	Basic Study Data for EPA Method (μg/1)	60
9-3	Sample Statistics: Sample 1	63
9-4	Sample Statistics: Sample 2	64
9-5	Sample Statistics: Sample 3	65
9-6	Sample Statistics: Sample 4	66
9-7	Sample Statistics: Sample 5	67
9-8	Sample Statistics: Sample 6	68
9-9	Results of Thompson's System	80
9-10	Results of Youden's Ranking	82
12-1	Standard Methods for Chemical Analysis of Water: List of Approved Test Procedures	128

LIST OF ABBREVIATIONS AND SYMBOLS

(1) μ = True mean (The expected value of a population, X, $\mu = E\,[X]$.)

(2) σ^2 = True variance (The expected value of the square of the difference between X and μ, $s^2 = E\,[(X-\mu)^2]$.)

(3) $\bar{X}$ = Sample mean ($\bar{X} = \sum_{n=1}^{n} \frac{1}{n} X_i$, where X_i, i = 1, 2,...,n are the results, and n is the number of results.)

(4) M = Median (Halfway point in the results when they have been arranged in order of magnitude (the middle result of an odd number of results, or the average of the middle two for an even number).)

(5) Accuracy (The correctness of a measurement, or the degree of correspondence between the results and the true value (actual amount added).)

(6) Precision (The reproducibility of sample results or the degree of agreement among the results.)

(7) E_m = Mean Error (The average difference with regard to sign between the results and the true value. Equivalently, the difference between the mean of the results and the true value (T.V.). E_m, Mean error = $\bar{X}$ - T.V.)

(8) E_r = Relative Error (The mean error expressed as a percentage of the true value. E_r, Relative error = $\frac{\bar{X} - \text{T.V.}}{\text{T.V.}} \times 100$)

(9) S^2 = Sample variance (sum of squared differences between measurements and sample mean, $\bar{X}$, divided by n-1, where n is the number of results.

$$S^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n - 1}$$

(10) S = Sample standard deviation (the square root of sample variance.)

(11) SD_r = Relative standard deviation (also called coefficient of variation; sample standard deviation normalized by the sample mean, $SD_r = \frac{S}{\overline{X}} \times 100$)

(12) R = Range (the difference between the largest and smallest results in the measurements.)

(13) t = Student's t distribution ($t = \sqrt{n}\,(\overline{x} - \mu)/S$)

(14) UCL = Upper confidence limit (the limit below which the true mean, μ, will lie with probability $1 - \alpha$, where α is the probability that the UCL does not bound the true mean. $UCL = \overline{X} + (t_{\frac{\alpha}{2}}\, S/N)$, where $t_{\frac{\alpha}{2}}$ is the upper $\frac{\alpha}{2}$ point of student's t - distribution.)

(15) LCL = Lower confidence limit (the lower counterpart of UCL, $LCL = \overline{X} - (t_{\frac{\alpha}{2}}\, S/N)$.)

(16) TL = Tolerance limits (limits within which one can state with proportion P of the entire population will lie. The upper and lower tolerance limits are given by $\overline{X} \pm Ks$, where K is the factor for two-sided tolerance limits for normal populations. The value of K depends upon the chosen values of γ and P.)

ACKNOWLEDGMENTS

The authors wish to convey their appreciation to Mr. Terry Covert, Project Officer, as well as Mr. Paul Britton and Mr. Edward Berg, all of the Environmental Monitoring and Support Laboratory, Cincinnati, Ohio, who have worked closely with us in bringing this project to a successful completion.

We wish to acknowledge the invaluable assistance furnished during our field investigation by Mr. James H. Finger, EPA, Region IV, Athens, Georgia; Mr. David Payne, EPA, Region V, Chicago, Illinois; Mr. Thompson, EPA-NERC, Research Triangle Park, North Carolina; and Mr. William Kelley, National Institute of Occupational Safety and Health, Cincinnati, Ohio.

The authors express their gratitude to the many Federal, state, and private laboratories whose advice and recommendations are incorporated into this project.

SECTION I

INTRODUCTION

NATURE OF THE PROBLEM

The role of the analytical laboratory is to provide qualitative and quantitative data that accurately describe the characteristics or the concentration of constituents in the sample submitted to the laboratory.

On the basis of the laboratory data, far-reaching decisions are often made. Water quality standards are set to establish satisfactory conditions for a given water use. Legal action is required by pollution control authorities when laboratory results indicate a violation of the standard. In wastewater analyses, the laboratory data define the treatment plant influent, the effectiveness of the treatment process, and the final load imposed upon the receiving water resources. Decisions on process changes, plant modification, or even the construction of a new facility may be based upon the results of laboratory analyses. The value and progress of research and development efforts depend, to a large measure, upon the validity of the laboratory results. In many cases, the protection of public health and the preservation of the nation's environmental resources are dependent upon the accuracy of laboratory analyses.

Because of the importance of laboratory analyses and the resulting actions which they produce, a program to insure the reliability of data is essential. An established routine control program applied to every analytical test is important in assuring the reliability of the final results. Furthermore, it is critical that analytical results between individual laboratories be accurate and precise. The additional variance between laboratories requires an established interlaboratory testing program to monitor and control individual laboratory performance. Once this performance is established as acceptable, then comparison of analytical results between laboratories can be meaningful and significant. Standardization of methods between cooperating laboratories is important in order to remove the methodology as a variable in comparison or joint use of data between laboratories. This is particularly important when laboratories are providing data to a common data bank or when several laboratories are cooperating in joint field surveys.

Under the charter of the U.S. Environmental Protection Agency, (EPA), the Office of Research and Development coordinates the collection of water quality data to determine compliance with water quality standards, to provide information for planning of water resources development, to determine the effectiveness of pollution abatement procedures, and to assist in research activities. To a large extent, the success of the EPA pollution control program rests upon the reliability of the information provided by the data collection activities.

The Environmental Monitoring and Support Laboratory (EMSL), Cincinnati, is responsible for insuring the reliability of physical, chemical, biological, and microbiological data gathered in water treatment and wastewater pollution control activities of the EPA.

The Quality Assurance Branch, Environmental Monitoring and Support Laboratory, Cincinnati, presently conducts formal interlaboratory studies among EPA laboratories to evaluate methods selected by EPA for its method manuals. Other federal, state, university and industrial laboratories are accomodated in these round-robin studies on a voluntary basis. The studies carry deadlines and conclude with reports distributed to all participants. Reference samples are also furnished without charge to interested governmental, industrial, commercial, and private laboratories for their within-laboratory quality control programs. However, there is no certification or other formal evaluation function resulting from their use.

Presently the EPA has no system for conducting interlaboratory tests to confirm laboratory proficiency. In the absence of such a system, certain doubts are raised as to the validity of the results reported by the Agency. Variances between laboratories are sources of errors which may have significant effects on the validity of the final data results.

Laboratories cooperating in joint survey programs or those providing results to a common data bank, such as STORET*, must maintain acceptable quality control to insure that analytical results between laboratories are in good agreement of accuracy and precision. The variance between laboratories must be maintained to an acceptable minimum if the final results are to be valid.

Because of the importance of water pollution data and the resulting actions they produce, it is essential that a dynamic system be developed and implemented by the EPA to conduct interlaboratory testing for evaluation of water and wastewater quality data and the laboratories producing these data.

---

*STORET is the acronym used to identify the computer-oriented EPA Water Quality Information System for STOrage and RETrieval of data and information.

OBJECTIVE

The objective of this program is to provide an interlaboratory testing program that will be one element of EPA's quality control evaluation system to be used for objectively evaluating the ability of an environmental laboratory under routine conditions, to analyze samples containing unidentified constituents in varying quantities, and to produce results that have the desired precision and accuracy for making valid decisions.

SCOPE OF WORK

To achieve this objective efficiently, this program has been divided into two phases. Phase I involved the investigation of existing interlaboratory testing programs using literature search and review followed by field investigation of Federal, State, and private laboratories. The data were analyzed and a preliminary program prepared.

Phase II consisted of final program development and a detailed program plan to be tested for functionality following the development of a program specification and method for testing.

SECTION II

SUMMARY

A dynamic system is developed for evaluating water pollution data and the laboratories which produce these data. The system consists of a plan for the design and implementation of an interlaboratory test program. A pilot test program is included to evaluate and to verify the complete program.

Investigation of interlaboratory tests conducted in the past has identified deficiencies in their design and in the procedures by which they were conducted. These conclusions are listed in Section III, and the recommendations which follow from them are listed in Section IV. The conclusions and recommendations are supported by an extensive literature review (Section V) of previous interlaboratory tests and their methods for experimental design and for test data analysis. Additionally, 18 EPA, State and private laboratory agencies were visited. Obtained by questionnaire and by personal interview, the comments, critiques, and suggestions of these agencies have served to identify major difficulties and deficiencies in interlaboratory test programs generally, and some specific causes of their failure to yield conclusive analytical and proficiency data. These field investigations are described in Section VI.

The interlaboratory test program developed in this study is presented in Section VII. The functions and responsibilities of each agency, namely, the cognizant EPA offices, the Interlaboratory Test Program Manager, and the participating laboratories are defined. Analytical methods and statistical procedures are specified for sample preparation, for data analysis, and for proficiency evaluation.

A pilot test program of limited scope, discussed in Section XI, is developed to test and to validate the experimental design and statistical analysis methods which have been selected.

Finally, a list of reference literature and publications is presented. This source material is extensive, and excerpts from it have been widely used in the body of the report. The contributions of these authors to the field of interlaboratory testing is hereby acknowledged.

SECTION III

CONCLUSIONS

1. Collaborative tests for methods development are well defined and are in wide use currently. Interlaboratory testing as a function of laboratory evaluation is under development and subject to many differing program objectives, design and statistical evaluations.

2. Interlaboratory test programs for proficiency evaluation must be carefully designed and implemented with adequate control procedures. Otherwise, the resulting data will be difficult to analyze and interpret. Regardless of the sophistication of the statistical analysis procedure, meaningful conclusions cannot be derived.

3. For proficiency evaluation to be effective, the number of participating laboratories should be as large as possible. This reduces the uncertainty associated with the test data statistics, and facilitates the differentiation among laboratories exhibiting nearly equal performance.

4. The interlaboratory test design must provide as large a number as possible of experimental data points, and these must be interrelatable (as, for example, in multiple Youden pairs). In this manner, the masking effects which result from gross errors inherent in many test methods can be minimized.

5. Prior to implementation, the interlaboratory test design should be validated, under controlled conditions, in one or two "reference" laboratories. This provides target level of performance, and permits ranking individual laboratories according to this target rather than according to the performance of their population at large.

6. The statistical methods employed in many prior interlaboratory tests are incomplete to the extent that they usually assume the data to be normally distributed yet fail to test and prove this assumption. Furthermore, they fail to derive confidence limits on sample means, standard deviations and relative ranking of laboratories.

7. Proficiency evaluation should not be limited to the analysis of individual laboratory test data, but should include evaluation of personnel qualifications, laboratory facilities and equipment, and in-house quality control standards.

8. Analytical results obtained in Method Study 7-Trace Metals, shows a wide variation in accuracy and bias errors, and seven of the sixteen laboratories performed in an unacceptable manner, based upon the proposed evaluation system.

9. When one or more laboratories fails to perform all tests as directed, or when results are reported as "less than x micrograms per liter", uniform statistics for each element and sample cannot be derived. However, even in these cases, an accurate assessment of individual laboratory performance can be obtained from as few as ten or twelve reported results of all the elements and concentrations to be tested.

10. Participating laboratories should be notified promptly of the test results and their individual levels of performance. Conclusions relating to individual performance should take into account data recording or transcription errors, equipment limitations, and procedural deficiencies.

SECTION IV

RECOMMENDATIONS

1. Proficiency evaluation programs should be closely integrated with the interlaboratory test design, to assure that the two functions are coherent and that the tests yield all information required for the evaluation.

2. The interlaboratory test design should include samples compounded from all five constituent groups, so that each laboratory may be evaluated with respect to all types of test and test procedures.

3. Laboratories selected for participation should be chosen from those which routinely and rigorously employ adequate quality control procedures. Otherwise, reported test data may be subject to large errors, and degrade the subsequent statistical analysis.

4. The statistical analysis should be as complete as possible so that timely and accurate program results may be reported to the participants.

5. Chemical samples to be used for laboratory proficiency tests should be compounded at concentration levels near those of prior method studies, or they should be subjected to tests of precision by two or three referee laboratories, to obtain target standard deviations at each concentration. These data are required to evaluate absolute (as opposed to relative) performance of each laboratory.

6. In order to standardize proficiency evaluation, the statistical analysis procedure adopted by the EPA should be published in "User's Manual" format for use by any interlaboratory test program manager involved in water quality measurements.

SECTION V

LITERATURE SURVEY

SURVEY OF EXISTING SYSTEMS AND METHODS (REF. 1-89)

The literature survey being conducted has encompassed three major subjects -- the evaluations of laboratory methods in the environmental field (Analytical Reference Service Reports), the reports concerning the laboratory accreditation program for industrial hygiene or environmental health laboratories, and the publications in the field of statistical methods of collaborative experiments. The results of the literature survey are summarized briefly in the following:

EVALUATION OF LABORATORY METHODS (REF. 1-36)

Interlaboratory test programs for evaluating analytical methods, such as those conducted by ASTM, AOAC, and IEPA, have been under development for more than three decades. The use of these programs for rating individual laboratory performance is relatively new. Consequently, most of the literature is primarily concerned with interlaboratory methods evaluation programs. The objective of these reports is the exchange of information so that accurate and precise analytical procedures can be agreed upon and followed by the laboratories involved. (Ref. 2-13) Cited are several typical failures in methods evaluation programs which arise from combinations of the following:

- Interlaboratory studies are poorly designed statistically; are not optimized.
- Data from such programs are not analyzed to determine which probability distribution pertains; frequently a correct parametric analysis is applied to an incorrect (non-normal probability distribution) data set.
- Methods are not subjected to Youden's ruggedness test procedures before the interlaboratory test, as evidenced by the gross disparity of data frequently obtained in such tests.

These reports have covered mainly the physics of water and the testing of the quality and pollution of water. The results of the tests by the various laboratories are tabulated and plotted as bar graphs with statistical quantities listed (mean, standard deviation, confidence limits, number of outliers, etc.). Not all reports give a complete statistical assessment about the precision and accuracy of the laboratories involved. In addition to the statistical inadequacies, there are several practical pitfalls that are as applicable today as they were in 1959 when Pierson and Fay (Ref. 14) identified them to be:

- Benefits derivable from the program not fully understood
- Chairman not fully qualified for the task and unaware of some of the requirements
- Objectives not clearly stated and understood
- Improper selection, preparation, or packaging of samples
- Inadequate written instructions from the chairman to the participants
- Inadequate statistical design
- Participating laboratories are not adequately instructed about the method prior to participation. Typically, where initial practice samples are supplied, there is an insufficient number to provide adequate experience prior to analyzing the test samples.
- The number of replicates required is frequently inadequate to determine the intralaboratory errors.

EVALUATION OF LABORATORY PROFICIENCY

Proficiency evaluation and certification programs for a variety of specialized analytical laboratories are presently in being or proposed by agencies of the Federal Government.

Typically, these programs contain three elements:

- Documentation submitted by laboratories

  Personnel qualification and duties
  Quality assurance program
  Standard analytical methods specified
  Facilities and equipment
  Records maintained

- Site visits

  Periodic visits are made to the laboratory by specialists who review the documentation and records and observe laboratory personnel performing analyses.

- Proficiency testing

  Laboratories participate in interlaboratory test programs on a regular basis. Current State-of-the-art is best described in documentation developed in promulgating these programs.

Proficiency evaluation and/or certification programs include:

U.S. Department of Health, Education and Welfare - Public Health Service

- Center for Disease Control - Atlanta, Georgia

  A formal program for proficiency testing and accreditation of clinical laboratories has been in force under the auspices of the Clinical Laboratories Improvement Act of 1967. In this program the interlaboratory test results are divided into three ranges. The laboratories whose results lie within the first and narrowest range are given a score of 3. The laboratories in the second narrowest range are scored 2; the laboratories in the third and the widest range are scored 1. Finally, the ones outside the widest range are given -1. The passing score is 1. The laboratories which fail are warned and asked to correct their measuring procedures. The limits used in determining the three ranges are based on three factors: (1) the central 95% of all laboratories under test, (2) reference lab values representing the true values, and (3) the clinical requirement, a percentage of the median of reference lab values. This program also adopts the histogram and X^2 tests as a two-level technique for normality. To test the significant difference among the laboratories, the program uses a short cut method in the analysis of variance.

- Food and Drug Administration - Cincinnati, Ohio

  Under the Grade "A" Pasteurized Milk Ordnance (1965) the Food and Drug Administration conducts a performance evaluation and certification program for state central milk laboratories who in turn certify official, commercial, and dairy-industry laboratories in the individual states. The approval of the milk laboratories is based on testing done twice a year, which includes two kinds of testing programs:

(a) Laboratory Survey Program (inspections of facilities, procedures, results, and records), (b) Split Sample Program (a minimum of 10, preferably 12, split samples being analyzed by each laboratory to show their accuracy as well as their precision). The statistical evaluation method includes the following steps:

(1) take the log of the viable counts and assume log normality.

(2) calculate the average as the estimated mean then reject the counts beyond 3σ range (σ is assumed known).

(3) recompute the mean and the 1.3σ range, note the laboratories that are outside the 1.3σ range (75% of the normal samples).

Presumably the laboratories that are consistently outside the 1.3σ range should be informed of their deficiencies.

- National Institute for Occupational Safety and Health (NIOSH) - Cincinnati, Ohio

  NIOSH is sponsoring a program being developed by the American Industrial Hygiene Association for accreditation of Industrial Hygiene Laboratories. According to the agreement between AIHA and NIOSH (National Institute for Occupational Safety and Health), the accreditation program calls for the laboratories to participate in the PAT 12, 13 (Proficiency Analytical Testing) Program so that the standards of the testing techniques are met. As a key element in the accreditation program, PAT program itself has been under study constantly; for example, NIOSH is developing a parametric testing method assuming log normality versus the ranking tests now in use.

  In addition, this agency is actively conducting programs for improving laboratory quality assurance programs and statistical methods for objectively measuring laboratory proficiency utilizing interlaboratory testing. NIOSH has published comprehensive information pertaining to analytical laboratory operation and quality control procedures.

U.S. Environmental Protection Agency

- Pesticides and Toxic Substances Effects Laboratory National Environmental Research Center, Research Triangle Park, North Carolina

  This EPA laboratory is responsible for coordinating a quality assurance program for 74 heterogeneous laboratory entities, which includes EPA, State and private laboratories that perform environmental pesticide analysis.

The statistical testing samples (or check samples) are distributed to the laboratories and the results are treated in the following steps:

(1) Compute the 95% range from the results and reject the labs with results outside the 95% range.

(2) Recompute the mean and standard deviation after the rejections and compute the relative standard deviation, which gives an indication of the overall accuracy and precision of the laboratories as a whole. (% total error).

(3) Rank the laboratory performance by a 200 point system - 100 points assigned to full identification and 100 points assigned to complete quantification.

(4) The laboratories scoring between 150 and 190 are taken as the ones with some definite problems to be resolved. The labs scoring below 150 should be advised to suspend all routine work pending the resolution of some very serious problems in measurement.

- EPA - Region V
  Central Regional Laboratory
  Chicago, Illinois and
  International Joint Commission
  Windsor, Ontario

The Upper Lakes Reference Group has established a program in determining the accuracy and the confidence which can be placed on the analytical data being produced by laboratories operating under different jurisdictions.

The interlaboratory testing program includes seventeen laboratories that analyze split samples utilizing a variety of "standard" methods. Further compounding the problem is the extremely low contamination levels compared to typical rivers and harbors samples.

Data are reported using Youden's graphical technique in addition to the standard statistical evaluations. "True" value of each sample is also calculated.

In addition to the Federal laboratory evaluation programs, several states have similar programs.

Typical of these are the programs of California and Illinois.

State of California, Health and Welfare Agency, Department of Health; Berkeley, California

California has an ongoing program for the certification of water laboratories. In 1974, the program was redirected toward more frequent recertification based on higher standards. Laboratory status is subject to change to reflect level of performance and technical facilities.

Interlaboratory testing is an integral part of the evaluation program. Laboratories with major deviations and/or omissions which are not correctable within a reasonable time will not be recertified to the proper authorities nor will their test data be accepted by:

California Department of Health
County and City Health Departments
California Water Resources Control Board
California Regional Water Quality Control Boards
Environmental Protection Agency - National Pollutant Discharge Elimination System
California Department of Fish and Game

Illinois Environmental Protection Agency (IEPA), Springfiled, Illinois

The three laboratories in the Illinois EPA system conduct interlaboratory testing in order to validate data produced and confirm the overall quality assurance program in operation.

This agency uses a formal quality control manual that was developed jointly by the three laboratories in the system using the following procedure:

(1) Analysts write each individual procedure

(2) Procedures are verified by test

(3) Signed by analyst authorizing procedure

(4) Approved by each laboratory after testing

(5) Procedure published showing effective date

(6) Procedures are subject to revision as required, but each revision has effective date.

This program minimizes the laboratory errors due to unauthorized deviations from standard methods employed.

Twin Cities Round Robin Program, Minneapolis - St. Paul, Minnesota

This volunteer program is composed of governmental, independent and industrial laboratories involved in the analysis of water and waste waters. The purpose of this project is to conduct interlaboratory tests covering five major parameters; demand, nutrients, metals, minerals and special constituents, then determine the correlation of results between laboratories in order to validate each laboratories' in-process quality control program.

The statistical evaluation consists of:

(1) Preparation of test samples by Youden's nonreplicate technique, i.e., preparing two similar yet different samples to be analyzed by each laboratory only once.

(2) Compute the sums and differences of the results by the laboratories from which the precision and total error are estimated.

(3) Perform an F test to check the presence of inaccuracy due to laboratories. Make recommendations if the inaccuracy (also called systemic error) is indeed present.

Approximately 16 laboratories participate in the program.

METHODS FOR EVALUATING INTERLABORATORY TEST RESULTS

The use of interlaboratory testing programs to evaluate a group of hetrogeneous analytical laboratories presupposes that the laboratories are meeting the following standards:

(1) Quality Control Program operating

(2) Trained technicians

(3) Instrumentation suitable for test

(4) Calibration of instruments routine

(5) Standardized measuring procedure

However, if the standard methods of analysis are not free of determinate error, then the above lab-to-lab deviations from norms cannot be discriminated from errors inherent in the method.

The methods currently in practice for evaluating interlaboratory test results include the basic statistical techniques such as computation of mean, standard deviation, analysis of variance, and tests for normality, and the techniques for discerning and correcting determinate errors such as Youden's two-sample technique, ranking methods, McFarren's true value technique, etc. It should be emphasized that once the determinate errors are detected for a particular laboratory, steps may consist of (1) training of personnel, (2) recalibration of instruments, (3) use of blanks, correction factors, or standard compensation, and (4) improvement of instrument to lower detection limits.

In the development of a test method, it is an accepted practice to consider the following factors; sensitivity, uncertainty of "calibration curve, ruggedness, and total error." These factors plus other statistical techniques are delineated in the following subsections.

Basic Statistical Techniques (Ref. 37-43)

Before the discussion of the basic statistical techniques in use, it is necessary to first define the statistical terms endorsed by ANALYTICAL CHEMISTRY. (Note: suggested when results reported are suited for statistical treatment, based on 5 or more determinations):

(1) Series. A number of test results which possess common properties that identify them uniquely.

(2) Mean. The sum of a series of test results divided by the number in the series. Arithmetic mean is understood.

(3) Precision Data. Measurements which relate to the variation among the test results themselves; i.e., the scatter or dispersion of a series of test results, without assumption of any prior information.

The following measures apply:

- Variance. The sum of squares of deviations of the test results from the mean of the series after division by one less than the number of observations.
- Standard Deviation. The square root of the variance.
- Relative Standard Deviation. The standard deviation of a series of test results as a percentage of the mean of this series. This term is preferred over "coefficient of variation."

The statistical methods in current use can be found in the literature cited earlier that are the publications of on-going programs, textbooks, and handbooks. These sources contain recommended standard approaches to:

(1) Control charts for quality control

(2) Estimation of standard deviation (σ) from average range of measurement.

(3) Test for difference of sample mean $\overline{X}$ versus population mean $\overline{X}$ (both σ known and unknown).

(4) Tests for difference of sample variance versus population variance.

(5) Test for differences of two sample means ($\overline{X}_1$ & $\overline{X}_2$); (both σ known and unknown).

(6) Test of normality

(7) Analysis of variance

The order of listing is not significant; the nature of the test suggests the basis for selection, and the methods enumerated are, as stated earlier, in general use. However, a suspected pattern of approach is discerned, apparent in part by virtue of its limitations. Principally, the definition of distribution seems to be implicit rather than explicit. Although a test of normality is provided, specification of other methods implies an underlaying assumption of normality. Whether this is a weakness or not remains to be determined. Second, it appears that some criterion for sample data classification would be of utility: should the data be agregated and treated in toto, or would some stratification contribute substantially to the analysis. Third, means for identifying and evaluating sources and effects of errors in the data are not provided. Fourth, the information of control charts is rather limited, providing reasonable basis for surmising that errors of precision of accuracy would be difficult to detect in a timely manner. It is likely that other more effective means for identifying such trends can be devised. The application of the results in the analysis of variance to collaborative testing is elaborated in detail by Youden. As the number of samples or the number of laboratories available for certain tests is seldom large, and the assumption of normality may not always be valid, it may be necessary to use methods of analysis based upon order statistics or some other form of nonparametric statistic.

Youden's Two-Sample Technique

The Youden "Two-Sample" technique requires the preparation of two samples that are similar in nature and fairly close in concentration.[1] Each laboratory is asked to measure the concentration only once. The measurements are labeled X and Y and entered in a chart as shown below. Each point in the chart represents the results from one laboratory.

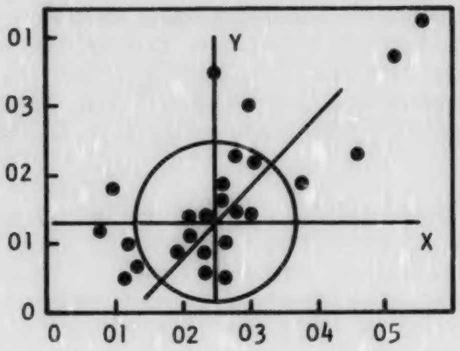

Figure 5-1. Percent of insoluble residue (illustration of the two-sample chart)

The pattern made by the points conclusively demonstrates the major role played by the differing error contributions.

Consider that random errors are really the cause of the scatter. Then the two determinations may err in being both low, both high, X low and Y high, and X high and Y low. As random errors are equally likely to be high or low (from the average), all four of the possible outcomes just enumerated should be equally likely. Thus, if a vertical line is drawn through the average of the X results and a horizontal line is drawn through the average of the Y results, the paper will be divided into four quadrants.

These quadrants correspond to the four outcomes, ++, --, -+, +-, just enumerated. If random errors are responsible for the scatter, the points corresponding to the laboratories should be divided equally among the four quadrants. In many hundred of cases no instance of such equal division has been found (unless the number of points is very small). The points are always found dominantly in two quadrants: the ++ or upper right quadrant and the -- or lower left quadrant. If a laboratory gets a result that is high (in reference to the consensus) with one material, it is almost sure to be similarly high with the other material. The same statement holds for low results. Generally, the points form an elliptical pattern with the major axis of the ellipse running diagonally at an angle of 45 degrees to the X-axis. Nearly always one or more points will be found far out along this diagonal clearly removed from the major elliptical cluster. The systematic error for the laboratory supplying the data for this point is evidently large in comparison with the other collaborators.

If random errors were vanishingly small, the amount high (or low) would be close to the same for both materials. The points would, therefore, hug the line closely, the ellipse becoming more and elongated. Indeed, the lengths of the perpendiculars drawn from the points to the 45-degree line are directly related to the random errors.

This technique has been adopted by many research laboratories and has also been used in interlaboratory testing programs. EPA Region V, Central Regional Laboratory, Chicago, Illinois; and Division Laboratories, California State Department of Public Health, Berkeley, California, use this technique extensively.[46]

The procedure of Youden was used, with some modifications, to evaluate the results from each laboratory. Youden's method not only permits the simultaneous evaluation of paired results, but also has the additional advantages of identifying results affected by systematic or random errors. The median value was determined for each constituent of each sample. These medians were used as In addition, to estimate overall precision, the standard deviation of the joint results was calculated according to the following formula: Estimate of Standard Deviation, SD,

$$SD = \sqrt{\sum_{i=1}^{n} \frac{d_i'^2}{2(n-1)}}$$

where $i = 1, 2, \cdots n$, n is the number of laboratories; $d'_i = d_i - \bar{d}$, d_i is the algebraic difference between results for sample 1 and sample 2 reported by laboratory i,

$$\bar{d} = \sum_{i=1}^{n} \frac{d_i}{n}$$

is the average difference between results for the two samples.

Table 5-1 summarizes the measures used to establish ranges of acceptable, questionable, and unacceptable performance.

The use of two water samples, analyzed for the same constituents, permits the application of an effective statistical technique. This procedure yields valuable data on laboratory performance that can be readily interpreted to the participants. Although a laboratory approval program that has high-quality performance as its goal can lean heavily on the use of reference samples, real laboratory improvement cannot be expected unless an appropriate adequate follow-up procedure is also instituted.

TABLE 5-1. MEDIAN VALUES OF SAMPLE CONSTITUENTS (TABLE 1 OF REFERENCE 46)

Constituent	Number of Laboratories Making Analyses	Median (mg/l)	
		Sample 1	Sample 2
Calcium	79	59.7	94.6
Magnesium	79	25.7	42.4
Sodium	73	29.0	106.5
Potassium	71	1.6	2.7
Chloride	91	33.8	168.3
Sulfate	78	42.8	111.0
Fluoride	81	0.84	0.44

Methods for Ranking Laboratories

In addition to the two-sample technique, there are several other techniques for the determination of performance levels of the laboratories. A simplified technique is the ranking procedure described by Youden[1], which involves the ranking of laboratory measurements according to actual data reported. For example, if A, B, C, laboratories report measurements of one sample as 1.5, 1.1, 1.8, then the ranks that the A, B, C laboratories receive will be 2, 1, 3 respectively. Lab A receiving rank 2 is considered most likely to have good performance. This ranking technique is most useful when a large number of labs and samples are involved. The following is an example when 10 labs and 5 samples are involved in a collaborative test where the data are arranged in a two-way classification scheme shown in the left half of Table 5-2.

TABLE 5-2. WATER-INSOLUBLE NITROGEN RESULTS (TABLE 6 OF REFERENCE 1)

	Results (%) For Samples					Ranked Results For Samples					
Column No.	1	2	3	4	5	1	2	3	4	5	Column Score
7	4.59	1.46	5.64	2.19	27.32	9	5.5	6	4	3	27.5
8	4.94	1.52	5.68	2.28	26.44	1	1	3	2	10	17
9	4.80	1.40	5.62	2.12	26.89	3.5	8.5	7.5	6.5	8	34
10	4.73	1.46	5.65	2.09	27.17	5	5.5	5	8	4	27.5
11	4.72	1.51	5.62	2.12	27.00	6.5	2.5	7.5	6.5	6	29
12	4.80	1.51	5.80	2.29	27.48	3.5	2.5	1	1	1	9[a]
13	4.45	1.40	5.45	2.07	27.02	10	8.5	10	9	5	42.5
15	4.72	1.50	5.58	2.27	26.76	6.5	4	9	3	9	31.5
16	4.63	1.32	5.69	2.04	26.92	8	10	2	10	7	37
17	4.88	1.42	5.67	2.16	27.39	2	7	4	5	2	20

[a] Designates unusually low score.

The right half of the table shows the data replaced with rankings that have been assigned to the laboratories according to the amounts reported to the referee. The rank 1 is given to the largest amount, rank 2 to the next largest, and so on. When a tie occurs between two laboratories for the xth place, each laboratory is assigned the rank x + 1/2. In the case of a triple tie for the xth place, all three get the rank (x + 1). This keeps the sum of the ranks equal to n(n + 1)/2, when n is the number of laboratories.

Each laboratory receives a score equal to the sum of the ranks it received. For M materials, the smallest possible score is M and the largest possible score is nM. A laboratory that reports the highest amount for every one of the M materials gets the score of nM. Such a score is obviously associated with a laboratory that consistently gets high results, and the presumption is that this laboratory has a pronounced systematic error.

We need a quantitative measure to pass judgment on the scores. We wish to know how big (or how small) a score we can reasonably expect to happen by chance in the total absence of any systematic errors. The numbers 1 to n may be written on n cards, which are then shuffled to obtain a random order for the ranks. Repetitions of the shuffling process will produce a series of random rankings for the laboratories. The scores will tend to cluster around the value $M(n + 1)/2$. The statistical distribution of such scores has been tabulated. When a collaborative test yields scores in extreme regions, we conclude that a pronounced systematic error is present in the work of the laboratory with the extreme score. In the face of such convincing evidence, the laboratory concerned should be willing to make a thorough search for the source of the systematic error. The referee may decide, in view of the evidence, to set aside all the results from this laboratory. Collaboratory 12 in the Table has a score of 9 as a consequence of getting high values rather consistently. The allowable score limits for 10 laboratories and 5 materials are 11 and 44. (Ref. 1)

In extreme cases, most of the laboratories may get approximately the same ranking on each material so that the scores approximate the values M, 2M,..., nM. Obviously, this is an indictment of the analytical method; presumably it is either inadequately written or unacceptably sensitive to the various environments encountered in the various laboratories.

Another ranking technique is the one used by the Center for Disease Control at Atlanta, Georgia, Public Health Service, HEW, where three ranges are established, as described in Section 1.2 Scores of 3, 2, 1 are given the labs that have reported measurements lying respectively in narrowest, medium, and widest ranges. The labs which lie outside the widest range are given a negative score of -1. The laboratories ranked below 1 are warned and asked to correct their measuring procedure.

A third ranking procedure in practice is the one adopted by the Presticides and Toxic Substances Effects Laboratory, National Environmental Research Center, Research Triangle Park, N.C., EPA. The testing and ranking procedures are described also in Section 1.2. The ranking procedure is essentially based on three criteria:

1. Identification of all compounds present

2. Correct quantitative assessment

3. Avoidance of reporting compounds not present

A 200-point system is used in actual ranking - 100 points assigned to full identification and 100 points assigned to complete quantification." For example, a laboratory is asked to identify and measure 5 compounds possibly present in a sample which actually contains 4 compounds. The laboratory reports 3 correct identification, 1 incorrect, and 1 missing. Since there are two incorrect identifications (1 missing and 1 incorrect), 40 points are to be deducted $(40 - 2 \times \frac{100}{5})$ from the 100 points, 100-40=60. A "compound quantification point" system is used according to the following definition:

$$\text{Compound Quantification Pt.} = \text{Comp. pt. value} - \frac{\text{Formulation-Value Reported}}{\text{Standard Deviation}}$$

Use the same example as above. Since there are 4 compounds, each compound is assigned 25 points (out of 100 points). If F is the formulation value, R is the reported value, and S is the standard deviation, then the compound quantification point (CQP) for the first compound is:

$$CQP_1 = 25 - \frac{F_1 - R_1}{S_1}$$

The total CQP will be

$$CQP = \sum_{i=1}^{4} CQP_i$$

For instance, the formulation of Dieldrin is 20 pg/ul, yet the laboratory reports 50 pg/ul. The quantification point is

$$25 - \frac{(20-50)}{2.5} = 13$$

where 2.5 is the standard deviation.

The sum of the identification points and quantification points is the total score of the laboratory. The research center has set the following ranking standard:

> The laboratories scoring between 150 and 190 are taken as the ones with some definite problems to be resolved. The labs scoring below 150 should be advised to suspend all routine work pending the resolution of some very serious problems in measurement.

Determination of Acceptable Analytical Methods

In testing an analytical method, it is not a simple task to determine whether the method is an acceptable one. This is because the data collected are subject to precision and accuracy errors. This is an important problem to standard methods committees if their selection of methods is to be sensible and unbiased. E. McFarren at Analytical Reference Service, Bureau of Water Hygiene, Public Health Service, Cinn., Ohio,[24] has proposed a method for judging the acceptability of analytical methods, which is borne out by certain of the ARS Evaluation of Laboratory Methods Reports, as well as by a recent article by Devine and Partington (Ref. 23). Clearly, total error has a bearing upon the statistical analysis of test results. When the total error is large, the relative evaluation of laboratories becomes statistically difficult. Furthermore, many of the current standard test methods may not be acceptable for interlaboratory evaluation, unless the inherent errors of the methods have been previously evaluated by suitable "ruggedness" tests. In this case, allowance can be made for systematic errors in the method.

Obviously, both precision and accuracy (as defined in ANALYTICAL CHEMISTRY) must be considered in judging the acceptability of an analytical method. The difference being in the case of collaborative study, that the precision as calculated from the data collected by many laboratories will be somewhat larger because of differences in reagents, instrument calibrations, glassware calibrations, etc. These latter errors are also random errors but are in addition to the operator or laboratory random errors calculated when a series of test results are collected by only one operator in one laboratory.

The mean error, on the other hand, as calculated for a series of test results from many laboratories may not bear any relationship to that calculated for a series of test results from one laboratory. The latter may represent either the method bias, the laboratory bias, or both. The former, however, since it is an average of the bias from many laboratories, presumably more truly represents only the method bias (accuracy).

Using slightly redefined terms for the precision and accuracy of collaborative data, it is possible by means of suitable statistical tests, such as the F test and the t-test to determine whether there is a significant difference in either the precision or the accuracy of two methods. If there is a significant difference, then the method that is either more precise or more accurate is, presumably, the better method. The terms slightly redefined are:[22]

1. Mean error - the difference between the average of a series of test results and the true result
2. Total error - (absolute value of mean error + 2 x standard deviation)/(true value in percent

The term standard deviation is the regular definition; namely, the square root of the variance. For example, let us assume that the following set of data was collected for two difference methods (Table 5-3).

TABLE 5-3. DATA FOR TWO DIFFERENCE METHODS

Method	Number of Results	Mean	Mean error	Std. dev.	Rel. error	Relative standard deviation
A	25	1.10	+0.10	0.05	10.0	4.5
B	25	0.90	-0.10	0.05	10.0	5.6

Application of the definition for total error gives:

A. $\frac{0.1 + 2(0.05)}{1.00}$ X 100 = 20% total error

B. $\frac{0.1 + 2(0.05)}{1.00}$ X 100 = 20% total error

which indicates that both methods are equally precise and accurate, as it should be. However, if one uses another definition for total error; for example, total error = relative error + 2 (rel. standard deviation), the results would be

A. 10 + 2(4.5) = 19% total error
B. 10 + 2(5.6) = 21% total error

and it appears that there is a difference in the two methods, when actually the methods are equally precise and accurate. This phenomenon occurs because, for A, the mean is greater than the true value (1.00), and for B, the mean is less than the true value. Consequently, one can conclude that the new definition by McFarren is more accurate in determining the acceptability of an analytical method. In addition, he also proposed to divide methods into at least three different classes; namely, methods that can be rated as excellent or highly satisfactory, methods that are acceptable provided no better method is available, and methods that are unacceptable. Since the experience of ARS has indicated that few methods will qualify even if a total error as large as 25% is permitted, those methods that do qualify might be considered acceptable only if no better method is available will have a much larger error, perhaps as great as 50%. Under these conditions, with reasoning similar to the above example, a relative standard deviation as large as 25% and a relative error as large as 45% would be acceptable. As can be seen, however, the permissible relative error is dependent on the size of the relative standard deviation and on the sum of the relative error plus two times the relative standard deviation not exceeding 50%. The third category then would be those methods that have a total error greater than 50% and that would be judged unacceptable.

In his paper, as a result of the application of the proposed criterion for judging the acceptability of analytical methods, atomic absorption spectrophotometry was found acceptable for the determination of zinc, chromium, copper, magnesium, manganese, iron, and silver but unacceptable for the determination of lead and cadmium. On the other hand, none of the pesticides studied could be determined satisfactorily by gas chromatography. Objective reevaluation with the proposed criterion of the methods, resulted in conclusions essentially in agreement with those previously determined subjectively.

Techniques for Testing Ruggedness of A Procedure

Once an analytical procedure is shown to be free of accuracy errors, it is then necessary to test whether the procedure will be rugged under routine conditions (both intralaboratory and inter-laboratory). This is to say, the procedure should be insensitive to a slight deviation from normal procedure. A technique for testing the ruggedness of an analytical procedure has been developed by Youden,[1] which is used in Reference 20, Section III. The details of such a technique are delineated as follows:

Let A, B, C, D, E, F, and G denote the nominal values for seven different factors that might influence the result if their nominal values are slightly changed. Let their alternative values be denoted by the corresponding lower case letters a, b, c, d, e, f, and g. Now the conditions for running a determination will be completely specified by writing down these seven letters, each letter being either capital or lower case. There are 2^7 or 128 different combinations that might be written out. Fortunately, it is possible to choose a subset of eight for these combinations that have an elegant balance between capital and lower case letters.

The particular set of combinations is shown in Table 5-4.

TABLE 5-4. EIGHT COMBINATIONS OF SEVEN FACTORS USED TO TEST RUGGEDNESS OF AN ANALYTICAL METHOD (TABLE 8 OF REFERENCE1)

	Combination or Detn No.							
Factor Value	1	2	3	4	5	6	7	8
A or a	A	A	A	A	a	a	a	a
B or b	B	B	b	b	B	B	b	b
C or c	C	c	C	c	C	c	C	c
D or d	D	D	d	d	d	d	D	D
E or e	E	e	E	e	e	E	e	E
F or f	F	f	f	F	F	f	f	F
G or g	G	g	g	G	g	G	G	g
Observed result	s	t	u	v	w	x	y	z

The table specified the values for the seven factors to be used while running eight determinations. The results for the analyses are designated by the letters s through z. To find whether changing factor A to a had an effect, we compare the average (s + t + u + v)/4 with the average w + x + y + z)/4. The table shows that determinations 1, 2, 3, and 4 were run with the factor at level A and determinations 5, 6, 7, and 8 with the factor at level a. Observe that this partition gives two groups of four determinations and that each group contains the other six factors twice at the capital level and twice at the lower case level. The effects of these factors, if present, consequently cancel out, leaving only the effect of change A to a.

Inspection of the table shows that whenever the eight determinations are split into two groups of four on the basis of one of the letters, all the other factors cancel out within each group. Every one of the factors is evaluated by all eight determinations. The effect of altering G to g, for example, is examined by comparing the average (s + v + x + y)/4, with the average of (t + U + w + z)/4.

Collect the seven average differences for A - a, B - b, ..., G - g, and list them in order of size. If one or two factors are having an effect, their differences will be substantially larger than the group of differences associated with the other factors. Indeed, this ranking is a direct guide to the method's sensitivity to modest alterations in the factors. Obviously, a useful method should not be affected by changes that will almost certainly be encountered between laboratories. If there is no outstanding difference, the most realistic measure of the analyticaly error is given by the seven differences obtained from the averages for capitals minus the average for corresponding lower case letters. Denote these seven differences by Da, Db, ..., Dg. To estimate the standard deviation, square the differences and take the square root of 2/7 the sum of their squares. To check the calculation, compute the standard deviation obtained from the eight results, s through z. Obtain the mean of the eight results. Square the eight differences from the mean, sum the squares, divide by 8 - 1, and take the square root. This estimate of the analytical error is realistic in that the sort of variation in operating conditions that will be encountered among several laboratories has been purposely created within the initiating laboratory. If the standard deviation so found is unsatisfactorily large, it is a foregone conclusion that the collaborative test should never be undertaken until a method has been subjected to the abuse described above and satisfactory results obtained in spite of the abuse. (Ref. 1 page 35).

The following is an example of the factors involved in a laboratory in measuring the percent of water in phosphoric acid samples. Table 5-5 gives the factors and eight measurements (Ref. 20, page 10.3.c).

TABLE 5-5. MEASUREMENT OF H_2O IN PHOSPHORIC ACID

Factor	No.	Letter	Value for Capital Letter	Value for Lower Case Letter
Amount of H_2O	1	A,a	Ca 2 ml	Ca 5 ml
Reaction Time	2	B,b	0 min	15 min
Distillation Rate	3	C,c	2 drops/sec	6 drops/sec
Distillation Time	4	D,d	90 min	45 min
N-heptane	5	E,e	210 ml	190 ml
Aniline	6	F,f	8 ml	12 ml
Reagent	7	G,g	New	Used

Measurement	s	t	u	v	w	x	y	z
	18.80%	20.58	19.90	18.03	19.50	19.16	19.88	19.85

One can proceed to calculate the ruggedness of the procedure to various factors. For instance, the sensitivity to a slight variation in reagent used is

$$\frac{s + v + x + y}{4} - \frac{t + u + w + z}{4} = 18.97 - 19.96 = -.99$$

Such computations are summarized as follows:

Condition Varied	Difference
Reagent	-0.99
Aniline	-0.83
Distillation Time	0.63
Amount of Water	-0.27
Distillation Rate	0.11
Reaction Time	0.09
N-Heptane	-0.07

From the summary, one can conclude that the reagent used, amount of aniline and distillation time exert greater effects on the analytical result than the other four factors. Therefore, the individual planning to use this test method must make the decision whether to redefine test procedure prior to using it in a proficiency testing program. This decision should be based on results of the test "pre-qualification" activity, which provide estimates of the accuracy and precision errors inherent in the method and for the particular samples.

If the differences, for example those shown above, are small compared with the estimates obtained from the pre-qualification, they can be considered acceptable. If they are not, and hence become significant contributors to bias (accuracy) error, then the method is not sufficiently rugged for interlaboratory evaluation.

SECTION VI

FIELD INVESTIGATION

Comprehensive meetings were held with AQC coordinators at selected EPA Regional offices, the EPA National Field Investigation Centers, NERC Laboratories as well as other Federal, State and private laboratories listed in Table 3-1.

The primary purpose of this survey was to:

- gather data from those agencies now conducting interlaboratory testing programs
- analyze and evaluate problem areas in existing programs
- obtain recommendations for alternative approaches

In order to obtain objective data in an orderly manner during the field investigation, a questionnaire was prepared and mailed to the agency to be visited at the time an appointment was made (Ref. 25). The questionnaire was divided into two sections:

Section I - Interlaboratory Test Programs

The preponderance of interlaboratory test programs, to date, have been concerned with methods development and evaluation. The questions in this secetion were intended to identify similarities and differences between interlaboratory test programs designed for methods evaluation and programs for rating individual laboratory proficiency.

Section II - Intralaboratory Quality Control Practices

Historically, only those laboratories with effective quality control programs have performed well in any interlaboratory test program. The questions in this section were designed to determine the level of quality control to be maintained in an analytical laboratory in order to maintain the levels of proficiency required for producing valid data to quantify water pollution.

The questionnaire is included as Appendix B. Due to the wide variation in purpose of the respondent laboratories, many questions were not applicable to all laboratories. However, the questions were successful in promoting candid discussions with the personnel. Each of these interviews produced new insights into the problems of developing interlaboratory programs for monitoring laboratory proficiency.

The deficiencies of prior interlaboratory test programs evaluated are listed in summary form on page 4. **Clearly,** no specific program has been deficient in all these respects, although most of them exhibit deficiencies in several areas.

The selection and preparation of test samples is a prominent shortcoming. For example, water samples distributed by the California Department of Health often contain constituent concentrations known to be below detection levels when the concentrated sample is diluted according to instruction. Any subject laboratory interested in retaining its certificate will be tempted to first test the concentrate, then the dilution using extraction techniques, and compare results.

Typical deficiencies reported during FMC's field investigation include the following:

1. Timely reporting of interlaboratory test results back to the participants seldoms occurs, resulting in reduced participation in voluntary programs.

2. Constructive advice on improving analytical capabilities is rarely furnished.

3. Many interlaboratory studies do not take into consideration the data parameters required to accomplish valid statistical analysis.

4. Analytical test procedures used in interlaboratory tests frequently result in gross disparity of data that can be attributed to the analytical procedure rather than to difference in personnel or instrument capability.

5. Samples are not received with adequate instructional material, properly preserved, and/or not in a quantity sufficient for analysis.

6. Concentration range of interlaboratory test not in normal range of routine testing conducted by laboratory. May be at or below detection limits of analytical instrument without special procedures, or may be more concentrated; i.e. Great Lakes Laboratories.

Other problem areas include:

1. Optimization of the frequency of interlaboratory tests for each of the major categories. For example Dr. Hall of the Center for Disease Control is attempting to decrease the frequency of some tests. Other agencies are limited to one or two testing

programs each year due to legislative requirements, level of funding, and/or difficulty in analyzing test data and preparing evaluation reports.

2. Analysis of test data and documentation of results appears to be a common problem, particularly when computer program development is concurrent with introduction of the interlaboratory proficiency testing program.

3. Some programs have reverted to Youden techniques in order to report results in a timely manner. In general manpower limitations of state and federal agencies preclude adequate followup with participating laboratories either from a shortage of experienced personnel or lack of jurisdiction.

4. Often unusual handling is given to the samples once they are identified as check samples by the analyst.

5. Trace analysis or analysis of low concentration parameters are not completely assessed by samples prepared by concentrates.

As a result of this investigation, no recommendations were made by the participants for alternative approaches to interlaboratory testing. However they did recommend that precautions be taken in data treatment. For example:

Under precautions to be observed in conducting interlaboratory proficiency testing, the following comments were made, "MDQARL'S Methods Studies do a good job in assessing methods but are insufficient in parameters and frequency to make any assessment of laboratory performance," and "MDQARL'S check samples are a great help in monitoring an in-house quality control program, but are not meant for proficiency testing in an interlaboratory program."

One major insight gained from the discussions during the field investigation was that no laboratory should waste its time and money in participating in methods development or performance evaluation interlaboratory test program until it has an effective intralaboratory quality control program in force.

An intralaboratory quality control program should be a documented program concerned with all aspects of a functional analytical program, i.e., adherence to sample preparation procedures, instrument calibration, etc.; precision and accuracy on each of group of samples; instrument stability over time (perhaps by use of check samples plus instrument use and repair logs); preparation and use of quality control reports such as computer files and/or quality control charts.

The objectives and procedures for conducting Methods Development and Proficiency Evaluation are individually unique and should not be confused when developing either program. The first is a impersonal technical evaluation where pride of authorship is about the only interpersonal relationship. On the other hand, proficiency testing is highly interpersonal with all the intangible values associated with the more highly developed procedures of rating individuals on their current ability and future potential.

TABLE 6-1. AGENCIES VISITED DURING FIELD INVESTIGATION

AGENCY	PRINCIPAL CONTACT
Environmental Protection Agency Environmental Monitoring and Support Laboratory Cincinnati, Ohio 45268	
Methods Development and Quality Assurance Research Laboratory 1014 Broadway Cincinnati, Ohio 45268	Mr. Dwight Ballinger, Director Mr. John Winter, Chief Quality Assurance Lab. Evaluation
Water Supply Research Laboratory Taft Laboratory 4676 Colombia Parkway Cincinnati, Ohio 45268	Earl McFarren, Chief Water Supply Division
National Environmental Research Center Research Triangel Park, North Carolina 27711	
Division of Atmospheric Surveillance Quality Control Branch	Mr. Seymour Hochheiser, Chief
Pesticides & Toxic Substances Effects Laboratory Chemistry Branch	Mr. J. F. Thompson, Chief
EPA Office of Enforcement and General Counsel	Lowell A. Van Den Berg, Deputy Director Dr. Richard Enderoux Carl R. Hirth
National Field Investigation Center 5555 Ridge Avenue Cincinnati, Ohio	
National Field Investigation Center Denver Federal Center Denver, Colorado 80225	Dr. T. O. Meiggs, Deputy Director
EPA Regional Offices Surveillance and Analysis Division	Regional Analytical Quality Control Coordinators
Region IV SE Environmental Research Laboratory College Station Road Athens, Georgia	James Finger, Quality Assurance Officer
Region V Central Regional Laboratory 1819 West Pershing Road Chicago, Illinois 60609	David A. Payne, Quality Assurance Officer
Region VII 26 Funston Road Kansas City, Kansas 66115	Dr. Harold G. Brown, Chief Laboratory Branch

TABLE 6-1. AGENCIES VISITED DURING FIELD INVESTIGATION

(Continued)

AGENCY	PRINCIPAL CONTACT
Region VIII Denver Federal Center Denver, Colorado 80225	John R. Tilstra, Quality Assurance Officer
Department of Health, Education and Welfare	
Public Health Service Center for Disease Control Atlanta, Georgia 30333	Charles T. Hall, Ph.D. Chief, Proficiency Testing Section Licensure & Proficiency Testing Branch
Public Health Service National Institute for Occupational Safety and Health 1014 Broadway Cincinnati, Ohio 45202	William D. Kelly, Deputy Director
Public Health Service Food and Drug Administration, Bureau of Foods Division of Microbiology, Taft Laboratory 4676 Colombia Parkway Cincinnati, Ohio 45226	James Leslie
National Bureau of Standards	
Office of Measurement Standards Gaithersburg, Maryland	Dr. Joseph M. Cameron, Chief
State Agencies:	
Ohio Environmental Protection Agency 1571 Perry Street Columbus, Ohio 43201	Dr. Edward E. Glod
Illinois Environmental Protection Agency 2200 Churchill Road Springfield, Illinois 62706	Arnold Westerhold
	David Schaeffer, Ph.D.
Private Activities	
American Council of Independent Laboratories 1725 K. Street, N.W. Washington, D.C. 20006	Mr. Robert Corning, Chairman Water Quality Sub-Committee Cedar Rapids, Iowa
Twin Cities Round Robin Program Minneapolis-St. Paul, Minnesota	Mr. William A. O'Connor SERCO Laboratories 2982 N. Cleveland Avenue Roseville, Minnesota 55113

SECTION VII

DATA ANALYSIS AND EVALUATION

OBJECTIVES OF DATA ANALYSIS AND EVALUATION

The objective of data evaluation in this study is to ascertain the accuracy and precision of the testing methods used by the various laboratories. After the evaluation, the data can be analyzed to determine the validity of various test methods and procedures. Specifically, data analysis and evaluation should serve two purposes:

A. Detection of error in the chemical determinations performed

B. Isolation and correction of the source of error

In principle, the "test sample" approach amounts to a calibration of the laboratories involved, very much akin to the "traceable" calibration of an instrument for physical measurement; if the organization and processes for a given determination in each laboratory are considered to be an "instrument" (Ref. 90-106).

GENERAL ANALYSIS AND EVALUATION PROCEDURES

The general procedures to be followed in data analysis and evaluation should be based on analysis of variance; namely, the one of analyzing the precision error (or the sum of squares between the labs). To aid the discussions on analysis and evaluation procedures, and laboratory training methods, one can first model the testing program as an information flow system where categorization of measurements with respect to experimental design can be visualized easily.

INFORMATION MODEL

Figure 7-1 presents the flow diagram of the suggested "baseline" information model. It is not intended, at this point, to be definitive, but to provide a framework into which various sample orderings may be placed. To establish this context, five stages of information are identified.

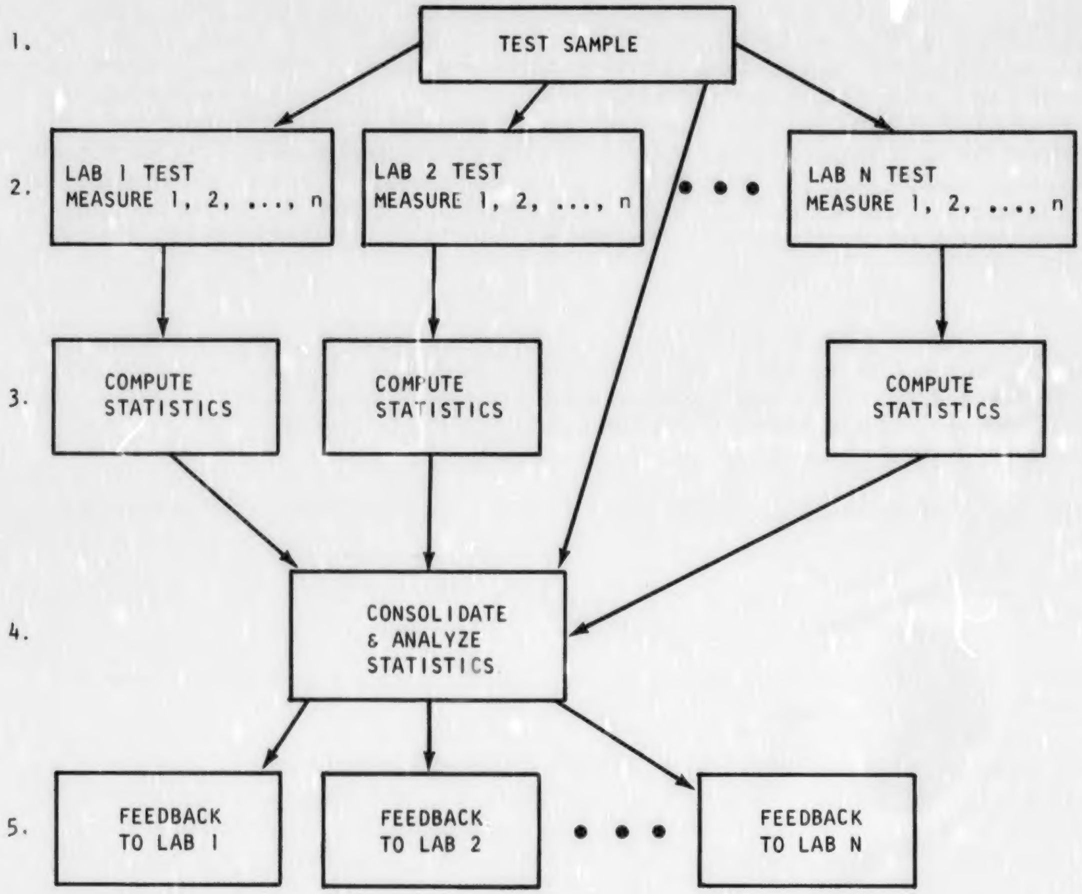
1.
TEST SAMPLE
2.
LAB 1 TEST
MEASURE 1, 2, ..., n
LAB 2 TEST
MEASURE 1, 2, ..., n
LAB N TEST
MEASURE 1, 2, ..., n
3.
COMPUTE STATISTICS
COMPUTE STATISTICS
COMPUTE STATISTICS
4.
CONSOLIDATE & ANALYZE STATISTICS
5.
FEEDBACK TO LAB 1
FEEDBACK TO LAB 2
FEEDBACK TO LAB N

Figure 7-1. Information flow model.

Stage 1. This is the test sample, containing the various ingredients to be determined in the test. The types and quantities of the ingredients are assumed to be unknown (at least to the individuals who will perform the test). It is further assumed that, within reasonable limits, the method by which the test sample is apportioned to the various laboratores does not affect the relative quantitative proportions of the ingredient.

Stage 2. This is the laboratory test; that is, the procedure applied to yield a determination of the types and quantities of the ingredients contained in the test sample. For each type of ingredient the sequence of measure; 1, 2, ..., n is recorded separately; this is the statistical sample, which should be treated as a member of the type/quantity population for succeeding analyses. (First caveat: if a particular type of ingredient is subjected to more than one technique of test determination, the results of different techniques must be treated as separate samples).

Stage 3. Each laboratory separately computes for each sample of measurements the statistics outlined on page 15 and following, at the minimum the steps described as the "mean" and "precision data." Where relevant information is available to the analyst, "accuracy data" may also be developed.

Stage 4. This is where the laboratory statistics are compared to:

a. The standard quantities of the sample, on a laboratory-by-laboratory basis.

b. The results elicited by each laboratory on a comparative basis.

Precision, accuracy, and test techniques are all subject to evaluation in this stage, and analytical methods are selected to provide the most effective means of comparison. Provided that historical data are available, trends in precision and accuracy with time may also be examined. This stage, in essence, is the quality control evaluation of the laboratories performing the test. Although indicated on the diagram as unique, information may pass through two or more substages, depending on the sequence of data consolidation. For example, if 100 labs produced test data samples, and as an intermediate step the statistics from sets of 10 labs were compared and 10 sets of consolidated results were provided to the final analysis. This intermediate analysis is of particular importance when two or more acceptable test methods have been specified in the test instructions. In this case, of the total population of participating laboratories, A group using method 1, B group using method 2, C group using method 3, etc., the group statistics are distinct from each other and caution is required in consolidating them.

Stage 5. This stage is a feedback to the participating laboratories. Principally, this feedback relates to the laboratory performance compared to that of the total population of laboratories. If the EPA is to elicit and to maintain a cooperative attitude among all laboratories, and this is desirable even if the intent of the test program is only compulsory participation as a condition of certification, then the participant must be given more than his bare "score". He is entitled to know his relative standing. If he is deficient, he should be told the nature of the deficiency, so that he may take appropriate action to rectify it. At the discretion of the EPA test program manager, he may even be provided with other samples to test and report on.

In summary, feedback should be regarded as a cooperative attempt by the lab and the analysis center to identify and eliminate causal factors for anomalies in test determinations.

SOURCES OF ERROR (Ref. 79-109)

As the objective of data analysis and evaluation is to detect, identify, and correct errors in an interlaboratory test, some examination of sources of errors is in order and some indication of their impact should be developed. For this purpose, the "test sample" is assumed to be standard; i.e., the type concentrations subject to determination can be known only within certain limits due to "Universe" error. Moreover, because of their rather specialized nature, the determinations considered are assumed to be for concentrations reasonably exceeding detection thresholds. The assumption establishes the general domain of labs that would undergo training, evaluation, and certification. The following diagram depicts the relationship among the errors.

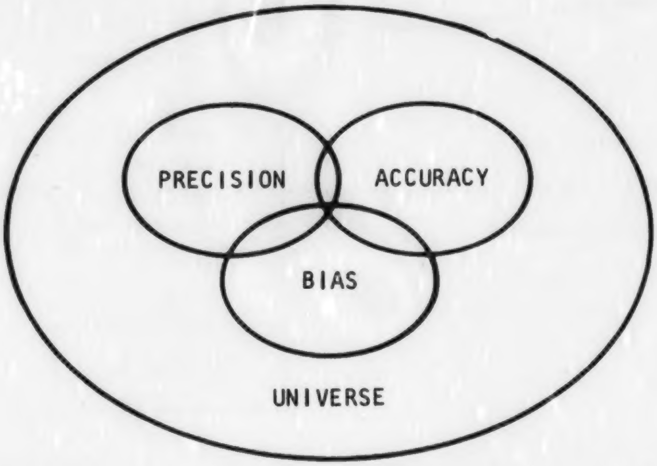

Figure 7-2. Error diagram.

Precision errors are errors randomly distributed about the mean. The expected value of their sum is equal to zero. Systematic (bias) errors are not distributed uniformly, and will yield a non-zero sum. Strictly speaking, bias and accuracy errors are synonymous; as the terms are used in this report, "accuracy" errors represent the performance of the individual analyst, while "bias" errors are characteristic of the laboratory itself and reflect the general laboratory environment. Both sources of errors result in a consistent displacement of test data from the true value.

The potential sources of error in a measurement imposed by the Universe are sketched as follows:

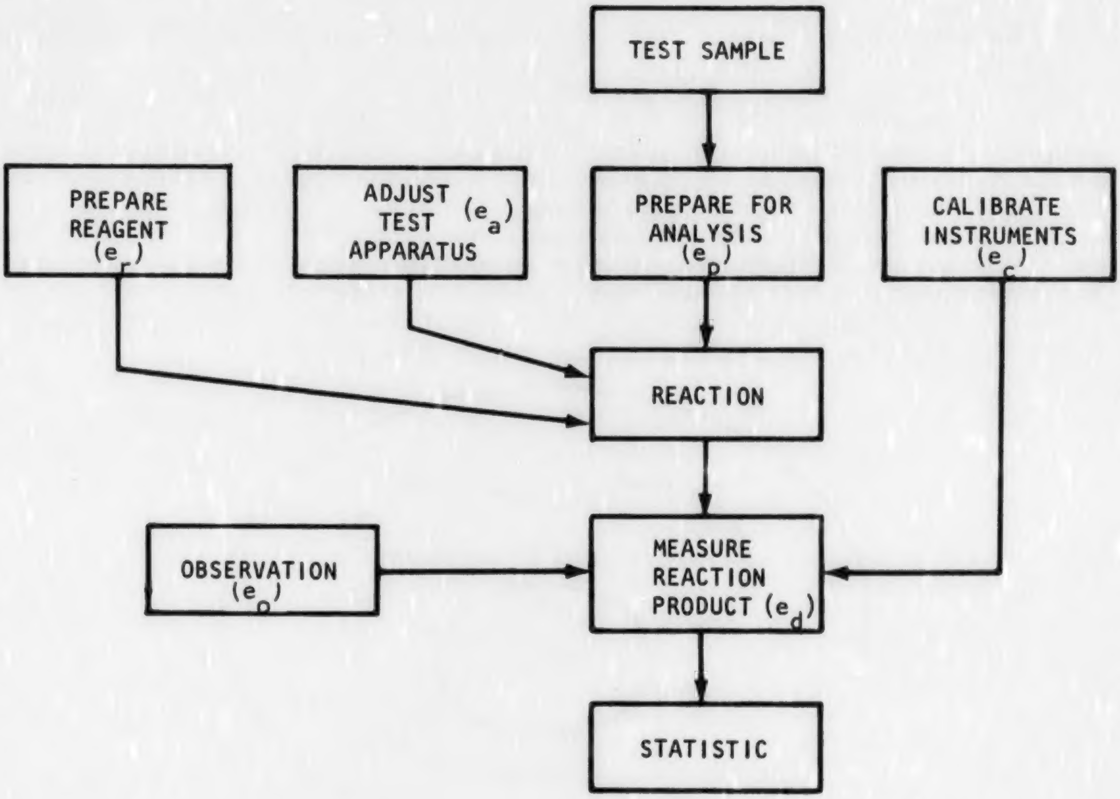
TEST SAMPLE
PREPARE REAGENT (e_r)
ADJUST TEST APPARATUS (e_a)
PREPARE FOR ANALYSIS (e_p)
CALIBRATE INSTRUMENTS (e_c)
REACTION
OBSERVATION (e_o)
MEASURE REACTION PRODUCT (e_d)
STATISTIC

Figure 7-3. Sources of Error in Measurement.

Each of the sources is discussed separately:

1. e_p - the error in preparing the test sample for analysis. This may involve dilution, the so-called "spiking" (which in essence shifts the precision parameters), or development of a more analyzable or measurable form.

2. e_r - the error in preparing chemical reagents. Whether the reagents are comparatively stable or their nature requires preparation immediately before development of a reaction, it is considered for this report that (typically) dilution or compounding introduce basic sources of error.

3. e_a - the error implicit in the installation and preparation of test apparatus. This factor is related to Youden's "ruggedness" criteria.

4. e_c - the error resulting from improper calibration of test instrumentation.

5. e_d - the error associated with round-off of data. Youden and others have suggested that the last digit be treated as having an inaccuracy of ± 1.

6. e_o - the error derived from reading and recording measurements by analyst; such things as misreading instruments, improperly manipulating apparatus, or inadvertently transposing digits fall into the category of human error.

To make the statistical analysis tractable, these sources are generally assumed to be independent of each other, the total error variance is then the sum of individual ones.

It should be noted that the human error listed above can be a predominating one. The accuracy of a recorded numerical value is subject to implicit limitation. This does not necessarily occur because of the "significant figures" consideration, though this appears to have been carelessly handled in a rather large fraction of the literature reviewed; typically squares or products of two-digit figures should result in four-digit figures or conversely roots or quotients result in corresponding reduction. Of somewhat more import here, however, is the simple act of interpreting what is recorded. Youden (and others) have suggested that the digit of least ordinal value (i.e. 2176) be treated as having an inaccuracy of plus or minus one (1) digit; for the number cited, falling between 2175-2177. If an interpolation of a measurement scale has been performed, the assumption is valid. On the other hand, if the reading is direct (as with digital readouts or such devices as analytical balances), the error must be assumed over a one-digit range; i.e., as previously cited, 2175.5-2176.5. (This conforms to the principle of "round-off"). The effect of the human element is not altogether separable. Such things as misreading instruments, improperly manipulating apparatus, or inadvertently transposing digits fall into the category of human error. Further, if a quantitative judgement is required, Weber's Law indicates that the detection threshold for differences is approximately 2.5% (relative). [Weber's law is a law in psychology, which has been thought to be the governing factor in human-initiated errors in reading measurements (Ref. 107)]. D. Meister has performed extensive studies on the effects of human errors in data collection procedures. The results of his studies indicate that a substantially high percentage of all equipment failures (20 to 80 percent) result from human error. He has also developed probabilistic theories to predict and measure human errors (Ref. 108, 109). Both Weber's law and Meister's studies can be used as references as to the extent of degradation on the accuracy of lab measurements caused by human errors.

SAMPLE SIZE REQUIREMENTS

The sample size required is generally a function of the parameters under estimate and the technique of sampling. Derivations have been carried out by constraining on the confidence intervals of the sample mean or sample variance, which result in expressions having the following form

$$n \geq f\ (\mu,\ \sigma^2,\ \bar{X},\ S^2)$$

where μ and σ^2 are the true mean and variance, and $\bar{X}$, S^2 are the corresponding sampled quantities. For example, by constraining on sample mean, one finds[39A]

$$n \geq \left(\frac{ts}{\bar{X}-M}\right)^2$$

By constraining on sample variance, one arrives at a similar expression but different from above mainly in the confidence interval. The latter constraint yields more stringent size requirement. Another significance about constraining on sample variance is that is is more meaningful in that even poorly designed tests which result in gross errors in the observed mean still may yield valid estimates of the measurement variances. It should be noted that the two constraints are equivalent when gross errors do not occur, that is, when the observed mean lies very close to the true mean.

OUTLIERS PROCESSING

The problem of outliers is a difficult one, especially in small sample cases where the only basis for rejecting outliers is the small number of samples which contains the suspected values. Youden recommends Dixon's approach which is to compare the gap between the outlier and the nearest value as a fraction of the total range and to reject the suspected outlier if the gap is greater than a certain fraction. (Ref. 1, page 30).

It should be noted that Dixon's approach is a nonparametric one which is generally not as powerful as a parametric approach. An alternate approach is not to reject the outliers as detected but to modify the values of the outliers. Such an amendment is quite attractive under the circumstances when the sample size is really small or one does not know the exact underlying distribution.

This amendment is called Winsorization 42, which can be demonstrated by an example: there are small samples, such as 7 labs involved in an interlaboratory study, reporting measurements as follows:

3.0, 4.2, 4.5, 4.7, 4.9, 5.1, 7.9

A Winsorization method with r=1 will make corrections on the extreme observations as:

$\underline{4.2}$, 4.2, 4.5, 4.7, 4.9, 5.1, $\underline{5.1}$

and compute the statistical parameters as usual,

$$\overline{X} = 4.67$$

$$S = 0.39$$

Without Winsorization, the results would be:

$$\overline{X} = 4.9$$

$$S = 1.49$$

The higher values of the mean and standard deviation, are due to the extreme observations, 3.0 and 7.9.

By a similar procedure, one can compare Winsoration with Dixon's Approach which dictates that the following ratios be checked.

$$\frac{7.9-5.1}{7.0-3.0} = 0.57$$

and

$$\frac{4.2-3.0}{7.9-3.0} = 0.24$$

According to Dixon (Table 8e in Reference 1), a ratio equal to 0.507 carries 5% risk of unjust rejection as an outlier. Since 0.57 is less than 5% risk one can reject 7.9 as an outlier. But, since 0.24 is less than 0.507, one cannot reject 3.0. The $\overline{X}$ and S are then computed based on 6 values,

3.0, 4.2, 4.5, 4.7, 4.9, 5.1

$$\overline{X} = 4.4$$

$$S = 0.75$$

This gives smaller $\overline{X}$ and larger S than those by Winsoration.

The important point in doing the Winsorization of data is that the effects of extreme observations are not completely thrown out, consequently, the danger of rejecting a lower estimate is greatly reduced. The efficiency of such a procedure has been shown to be quite high. Moreover, it also desensitizes the estimate to variations in the tails of the underlying distribution.

As stated earlier, a parametric outlier detection method is more powerful than a nonparametric one. For example, Grubbs' method is found to be more desirable by NIOSH since the log normality has been established for the data.[13] This method essentially uses t statistics, from which the maximum of the absolute t values is found and compared with an established table of critical values. Outliers are subsequently detected and rejected. For an example of this method, see Section 9.

The general procedure is illustrated in Figure 7-4, and a discussion of it is contained in Appendix C, Reference 4.

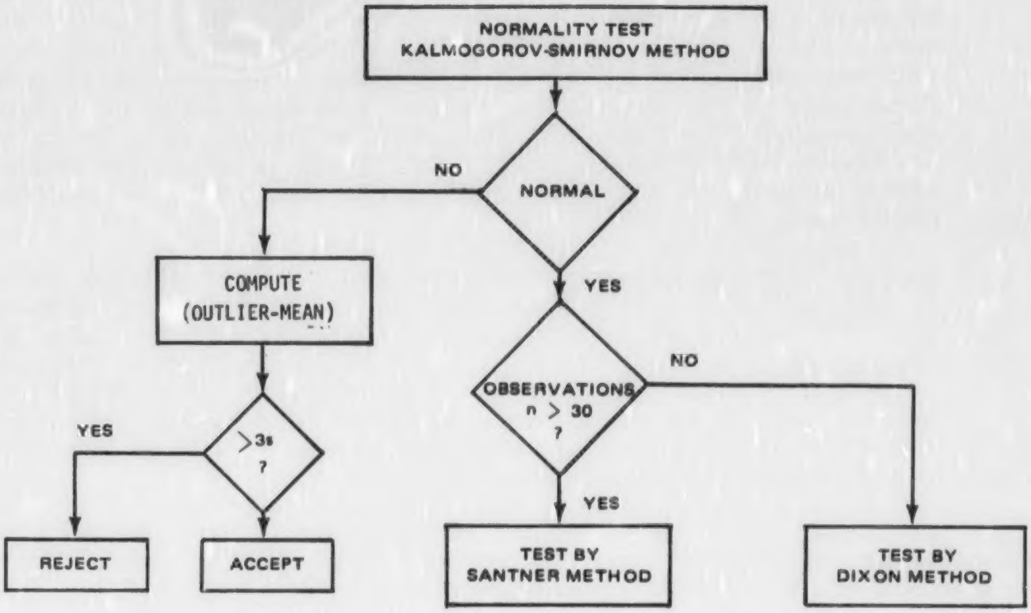
NORMALITY TEST
KALMOGOROV-SMIRNOV METHOD
NO
NORMAL
YES
COMPUTE
(OUTLIER-MEAN)
NO
OBSERVATIONS
n > 30
?
YES
>3s
?
YES
REJECT
ACCEPT
TEST BY
SANTNER METHOD
TEST BY
DIXON METHOD

Figure 7-4. Normality test and outliers treatment

METHODS FOR LABORATORY TRAINING AND DATA EVALUATION

Based on the literature survey and data management discussed so far, an outline of the two-phase program is presented here to conduct interlaboratory tests for water quality and effluent measurements. The first phase is a training program in which the laboratories involved are subjected to quality control training so that the lab errors including the precision error and accuracy error are minimized. The training procedure may consist of field visits (inspection) by qualified personnel, a survey of performance and a sample testing program.

A. Laboratory Visits - To be carried out by qualified inspectors to determine the quality of the equipment, procedures, results, and personnel involved in performing the experiments. Judgments should be made as to the degree of compliance with quality standards. Recommendations should be made for improvement where required. This survey should cover all the details of the laboratory facilities, measuring methods, techniques of recording, and personnel qualifications. The returns from the survey should be tabulated and analyzed to identify the problem areas as well as the differences and similarities among the labs. Recommendations should be made to resolve the differences and the problems.

B. Sample Testing Program - Formulated samples should be sent to participating labs for the identifications of compounds and the quantitative analysis. Statistical data analysis should include the evaluation of means, standard deviations, relative standard deviations (coefficient of variation) and percent total error. The relative performance rankings should also be shown to identify those laboratories where requirement for improvement is indicated.

This training program should be repeated twice a year, for example, to ascertain that appropriate quality and level are maintained by all the laboratories.

The procedures of data evaluation should include the following steps:

- Data Screening. This is a step to identify and reject extreme measurements (outliers) which can be done by Dixon's method for small sample case or by the 95% range method as used by Center or Disease Control and by Pesticides and Toxic Substances Effects Laboratory for large sample case. In addition, the Winsorization technique should be considered as a candidate method. If there is doubt about the normality of the population, one may apply histogram, X^2, testing, or Kolmogorov-Smirnov goodness-of-fit testing technique before the rejection of outliers is carried out (Ref. 3-8).

- Computation of Statistical Values. After the outliers are rejected, one can proceed to calculate the mean, standard deviation, percent total error, etc. The limits should also be calculated for various confidence levels.

- Ranking of Lab Performance. This step is done to provide an indication of the improvements for the laboratories with poor performance. In this step, a set of reference laboratories might serve as a yardstick for ranking performance. The reference laboratories should be ones known to have high-quality personnel and facilities, and long history of satisfactory performance. The performance ranking technique used by Pesticides and Toxic Substances Effects Lab and the ranking technique suggested by Youden are two possible candidates.

The procedures of these two-phase programs of lab training and data evaluation are summarized in Table 7-1.

TABLE 7-1. LAB TRAINING AND DATA EVALUATION

Lab Training	Data Evaluation
1. Lab visit	1. Data Screen
2. Lab survey	2. Statistical Computations
3. Sample testing	3. Performance Ranking
Existing Similar Programs	Existing Similar Programs
(i) PAT Program	(i) All the Programs in LAB TRAINING
(ii) USDA Milk Lab Program	(ii) EPA, Analytical Reference Service Reports
(iii) Twin Cities Round Robin Program	(iii) EPA Surveillance and Analysis Divisions (Georgia, Illinois, etc.)
(iv) Public Health Disease Control Lab	
(v) EPA, Research Triangle Park, PTSEL	
(vi) Training Program by Env. Health Facilities, Cincinnati, Ohio	

SECTION VIII

INTERLABORATORY TEST

PROGRAM PLAN

The interlaboratory test program for water quality and effluent measurements is intended to provide a method for the periodic assessment of the performance of the 22 EPA laboratories (and potentially 50 state laboratories) which routinely perform these measurements. The documents, procedures, and statistical methods which have been reviewed as a part of this study (see Sections 5 and 6 above), and information obtained in visits to and correspondence with laboratories engaged in water and waste water analysis, together form a well-defined background of experience and practice within which water quality tests have been performed during the past 20 years. In spite of the effort which has been applied to the conduct and analysis of these tests, inadequate emphasis has been directed toward interlaboratory testing sufficient in scope to yield valid conclusions regarding the natural environment, the degree and extent of local disturbances (industrial and agricultural), and the validity and consistency of analytical results produced by testing laboratories which measure these products.

The following paragraphs of this section describe the program and its elements in detail. In summary, the program consists of the following major activities:

1. Selection or Designation of Participants - The test program manager must determine which laboratories are suitable subjects for proficiency tests. Presumably, the proficiency tests are to be used as one criterion in initial certification and periodic recertification of federal, state and local governmental and private commercial laboratories which routinely perform water and waste water quality measurements.

2. Test Schedule - The schedule and frequency of tests, annually, semi-annually or quarterly, will influence the type and number of samples to be tested. If performed annually, the test program must cover representative elements and measurements of all five groups (Demand, Nutrients, Metals, Minerals, Special) If performed quarterly, the number of samples and measurements required for each test may be correspondingly reduced.

3. Selection and preparation of Samples - For reasons developed in Section IX below, sample element concentrations should be selected at levels for which Method Study results are available. This constraint arises from the need to have a prior statistical basis for evaluating absolute performance.

4. Preparation of Instructional Material - Participants must be instructed to perform all tests, and to report numerical values for each result, unless the element is reported as "not detected".

5. Mathematical Analysis - Laboratory proficiency will be measured as a function of the accuracy of each reported result. Methods are presented for combining the individual results reported by each laboratory, in order to assess its overall performance relative to other participating laboratories and relative to the standards shown to be achieveable by referee laboratories or Method Study results.

6. Report of Test Results - The test program manager must prepare and distribute a report of the test. This report will describe the overall test results, and it will contain an insert for each laboratory which identifies its performance. In the event that the laboratory performed poorly, the insert material will contain suggestions or instructions for improvement.

PROGRAM SCOPE

A test program of the magnitude defined herein involves the integrated activities of several organizations within the EPA, the test program management function, and the participating laboratories. Within EPA several ongoing programs are related to the interlaboratory tests, and these include separate studies on sample preparation and handling, interlaboratory evaluation protocol, laboratory certification, and the laboratory data storage and retrieval system (STORET). The interlaboratory test program management function is a control activity which involves many interfaces with other EPA organizations as well as with the participants being tested.

The interrelations among these activities are illustrated in Figure 6. Each activity is described briefly below, and the major items are discussed in greater detail in the following sections of this program plan. Heading numbers correspond to the block identifiers of Figure 8-1.

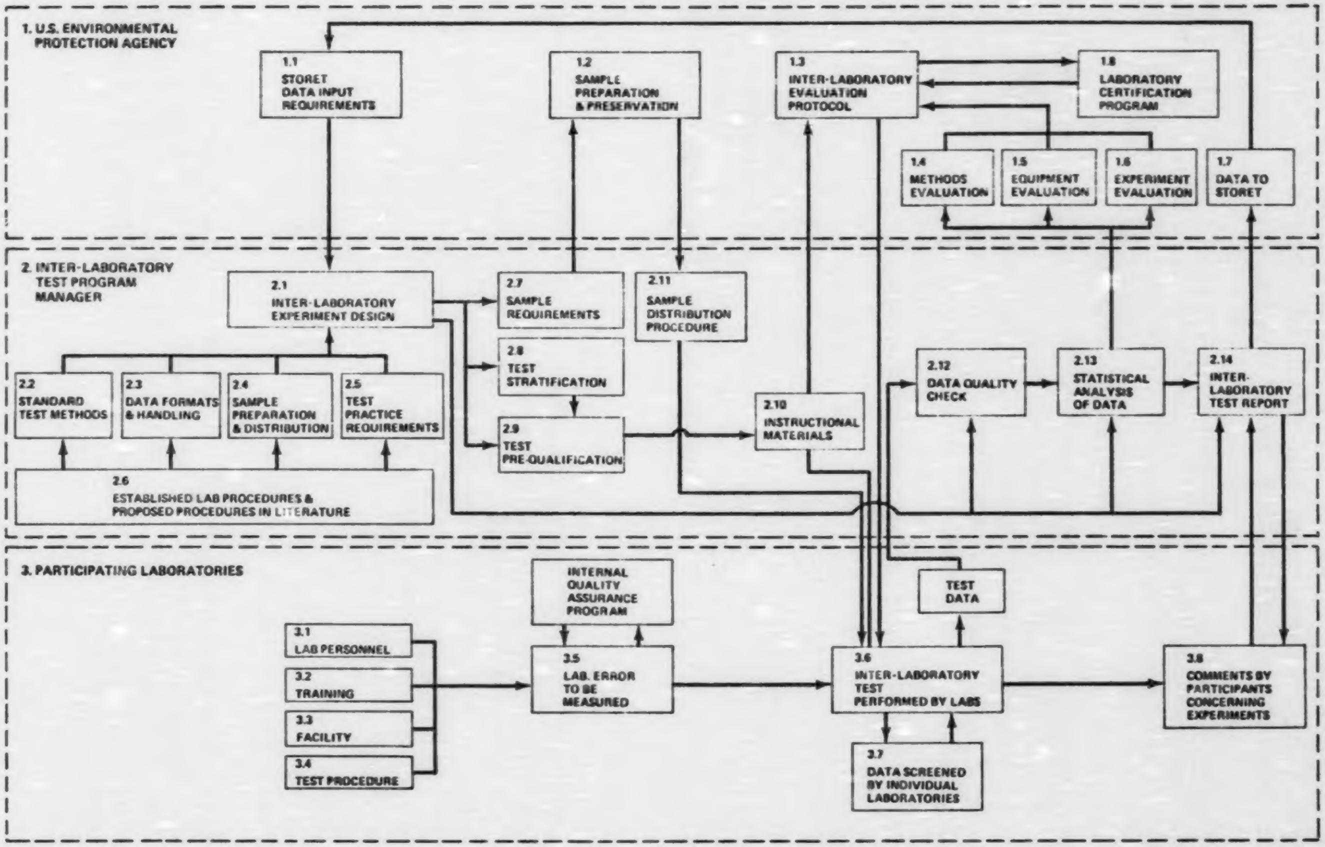
1. U.S. ENVIRONMENTAL PROTECTION AGENCY
1.1 STORET DATA INPUT REQUIREMENTS
1.2 SAMPLE PREPARATION & PRESERVATION
1.3 INTER-LABORATORY EVALUATION PROTOCOL
1.8 LABORATORY CERTIFICATION PROGRAM
1.4 METHODS EVALUATION
1.5 EQUIPMENT EVALUATION
1.6 EXPERIMENT EVALUATION
1.7 DATA TO STORET
2. INTER-LABORATORY TEST PROGRAM MANAGER
2.1 INTER-LABORATORY EXPERIMENT DESIGN
2.7 SAMPLE REQUIREMENTS
2.11 SAMPLE DISTRIBUTION PROCEDURE
2.8 TEST STRATIFICATION
2.9 TEST PRE-QUALIFICATION
2.10 INSTRUCTIONAL MATERIALS
2.12 DATA QUALITY CHECK
2.13 STATISTICAL ANALYSIS OF DATA
2.14 INTER-LABORATORY TEST REPORT
2.2 STANDARD TEST METHODS
2.3 DATA FORMATS & HANDLING
2.4 SAMPLE PREPARATION & DISTRIBUTION
2.5 TEST PRACTICE REQUIREMENTS
2.6 ESTABLISHED LAB PROCEDURES & PROPOSED PROCEDURES IN LITERATURE
3. PARTICIPATING LABORATORIES
INTERNAL QUALITY ASSURANCE PROGRAM
TEST DATA
3.1 LAB PERSONNEL
3.2 TRAINING
3.3 FACILITY
3.4 TEST PROCEDURE
3.5 LAB. ERROR TO BE MEASURED
3.6 INTER-LABORATORY TEST PERFORMED BY LABS
3.7 DATA SCREENED BY INDIVIDUAL LABORATORIES
3.8 COMMENTS BY PARTICIPANTS CONCERNING EXPERIMENTS

Figure 8-1. Inter-Laboratory Test Program

U.S. Environmental Protection Agency

STORET Data Input Requirements--
The interlaboratory test program shall be designed to accommodate the input data requirements of the STORET environmental data processing system. These requirements do not have direct impact upon the design and conduct of the interlaboratory test program, but the data recording, reporting and analysis formats and procedures should be constructed to provide a proper input format to the STORET system.

Sample Preparation--
The EPA test coordinators who are responsible for the preparation and distribution of test samples will be able to utilize the results of the Interlaboratory Experimental Design activity below to determine the quantities and constituents of samples which must be supplied for each test. These requirements will supplement the work separately being performed under EPA Contract No. 68-03-2075.

Interlaboratory Evaluation Protocol--
This activity, EPA RFP CI-74-0412, is intended to develop a uniform method for the evaluation of laboratory performance. The interlaboratory test program will provide the instructional material for each test (Instructional Materials), which will serve as a basis for evaluating the capability of any specific laboratory to perform environmental monitoring procedures.

Evaluation--
Three separate activities comprise the test program evaluaion function. The first of these, Methods Evaluation, consists of the ongoing monitoring and periodic revision of standard test methods and procedures. Similarly, laboratory equipment used in the tests is evaluated, from time to time to assess the capability of equipment in general use, and to investigate the capability and limitations of new equipment introduced into the field.

The third evaluation function is concerned with the experiment itself. The EPA will examine the overall performance of all participating laboratories, to determine that the levels expected from the analysis of the experimental design activity (Test Stratification and Test Prequalification) are achieved. If the participants uniformly fail to meet the expectation, then the test itself becomes suspect, and should be redesigned.

Data to STORET--
This function consists of the collection of preprocessing of interlaboratory test results, in a format suitable for input and retrieval in the STORET system.

Laboratory Certification--
In the event that the EPA implements a formal certification

program for environmental monitoring laboratories, the initial certification and periodic proficiency reviews may utilize the individual and collective results of the interlaboratory test program. Even if a formal program is not undertaken at this time, the Laboratory Evaluation Protocol (see Interlaboratory Evaluation Protocol) and the interlaboratory test results will provide a technical basis for establishing minimum performance criteria for future use.

Interlaboratory Test Program Activities

Interlaboratory Experiment Design--

The scope of this activity is primarily determined by the overall requirements for laboratory evaluation and certification, and the number and typesof tests and number of participants are functions of these factors. However, the detailed design shall be developed from technical criteria which reflect current practice and the inherent limitations of test methods and laboratory facilities.

Inputs to Experimental Design--

Current and proposed procedures constitute the main source of inputs to the design of interlaboratory tests. These include:

Interlaboratory Test Programs - conducted by EPA, USDA, Pat Program, PHS, and other agencies.

Data Formats and Handling - the mathematical and statistical techniques of data acquistion, reporting and analysis.

Sample Preparation and Distribution - including the determination of constituents and their concentrations, and handling by the laboratory.

Test Practice Requirement - as a part of instructional materials to familiarize laboratory personnel with the desired test procedure.

Sample Requirements--

As determined by the experimental design, requirements for samples for each test will be established. These will include the total quantity and number to be supplied to each participating laboratory (see Sample Preparation).

Test Stratification--

Some interlaboratory tests will require personnel and equipment capabilities beyond the scope of the typical commercial or local government laboratory, and will be restricted to a smaller number of laboratories, typically at the state and federal level. Because of the smaller number of participants, consideration must be given in the design itself (for example, a factorial design) to assure that statistics are properly defined. More general tests will involve a larger number of participants, and will provide a broader data base.

In both cases, effort will be directed toward minimizing the time and cost associated with each series of tests, and the consequent burden upon the participants.

Test Prequalification--

Each experiment should be prequalified by performance at a small number of well-qualified laboratories. This achieves two purposes. First, any deficiencies in the design itself can be identified before field experiments are undertaken. Second, the test results of these laboratories will serve as a target or baseline for the performance expected of the public or private laboratories to be evaluated.

Instructional Materials--

These materials will include a statement of the nature and objective of the test, a detailed procedure for handling and preparation of the test sample and necessary supplies, and a statement of special precautions or qualifications, if any.

They will also include forms and specifications for data recording, and for any mathematical operations which the laboratory shall follow using the raw data. Similarly, requirements for descriptive or commentary text, will also be specified.

The required scope of instructional material is typified by the protocols supplied to participants by the HEW Center for Disease Control, by the FDA milk testing program, and by EPA Region V Surveillance and Analysis Division.

Sample Distribution Procedure--

The interlaboratory test manager shall be responsible for specifying and controlling the preparation and distribution of samples for each test series, in coordination with the activity of Sample Preparation.

Data Quality Check--

When test data has been received from all participants, the data shall be reviewed as submitted to assess quality. Before performing statistical analysis, the test program manager shall examine and attempt to resolve apparent anomalies. The disposition of such results shall be based upon the manager's technical judgment as to their utility in the combined analysis.

Statistical Analysis of Data--

The techniques to be followed in statistical analysis, appropriate to the nature and size of the test, will follow from the test design (Interlaboratory Experiment Design)

These have been discussed in detail in Section 7 of this report by FMC, and are further elaborated below. Typical methods are contained in the Quality Control handbook published by NIOSH, and in the Industrial Hygiene Laboratory Accreditation program of the Center for Disease Control.

Report--

The final report for each test series will incorporate the data analysis and experiment analysis described above. It will also include commentary material submitted by the participants bearing on the test, the procedure, and the individual results.

Participating Laboratories

The evaluation of individual laboratory performance involves the assessment of all contributions to error in the test results. These include random and systematic components of accuracy, precision and laboratory bias. The major contribution to these errors are discussed below.

Personnel--

Personnel contributions to laboratory error include those which, at any level, may lead to or result in obtaining and reporting test data.

At the administrative level, these may include the misallocation of the personnel, instructions contrary to the required protocol, and errors in the processing of paperwork.

At the technical level, personnel errors include errors in procedure, improper use of materials and equipment, faulty interpretation of instrumentation and of visual test results, and mistakes in recording or manipulating test data.

Training--

Deficiencies in indoctrination and training, although they result in personnel errors, can be separately evaluated. They can be eliminated by proper briefing as to test objectives and procedural requirements, familiarization with test materials and equipment, and indoctrination into the required experimental protocol.

Facility--

Facility deficiencies include lack of required supplies and equipment, proper allocation of space, heating, lighting and ventilation, and any other environmental condition which may contribute to experimental error.

Test Procedure--

It is assumed that approved test procedures will be followed and that no fundamental errors are incorporated into them. However, apparent ambiguities may occasionally remain, which, although understood by a properly qualified operator, may lead to experimental error by less qualified personnel. If such deficiencies occur, they shall be identified and corrected by the evaluation of results for each experiment.

Laboratory Error--

Laboratory error is the composite of the preceeding four factors. The separate identification of random errors)precision) and systematic errors (accuracy and laboratory bias) is the primary objective of the interlaboratory test program. The frequency and magnitude of these errors can be at least partially controlled by an active Quality Assurance program operating within each participating laboratory.

Interlaboratory Test--

The test shall be performed in complete compliance with the prescribed protocol. However, the laboratories should be instructed to perform as nearly as possible in their normal practice. If more than one analyst participates in the test, individual records should be reported.

Screening of Data--

Errors in reported test results can be evaluated to some extent in the laboratory itself by means of data screening. While some errors may remain, such as failure to detect the presence of a potential constituent, or human error in interpreting test data and results, nevertheless each result should be examined for "reasonableness." If the result appears anomalous, then the laboratory should repeat the test to verify the finding, and report the difficulty to the test program manager.

PREPARATION AND DISTRIBUTION OF SAMPLES

The test samples used in this program will be constituted into each of the five groups as conventionally defined:

I Demand

II Nutrients

III Metals

IV Minterals

V Special

Three samples will be prepared for each group and the concentrations of each constituent shall be chosen to lie:

1. At or below detection limits unless special instrumentation and/or extraction processes are employed.

2. At or slightly above detection limits of a well-instrumented laboratory. Extraction may be required in some cases.

3. Near normal levels reported for surface waters. Interferences shall be passed in some cases.

The concentrations of each constituent shall also be selected to permit the mathematical analysis of laboratory results by Youden pairs and the other statistical methods described in Section 4 and 6 of this report.

Before the samples are distributed to all participants, six variants of each sample will be analyzed by a laboratory selected as a reference. The first two shall be prepared at 50 percent below the nominal value of each constituent, the second two at nominal values, and the third two at 50 percent above the nominal values. These variants may be constituted simply at different dilutions; however, the dilution shall be performed prior to sample distribution by the Interlaboratory Test Program Manager. The purpose of this "prequalification" is the assessment of ranges in accuracy and precision to be expected from the results reported later by all participants.

The Test Program Manager shall also prepare instructions for sample storage, distribution, and handling by the reference laboratory and by all participants.

INSTRUCTIONAL MATERIAL

Instructional material used in this program shall consist of three types. First, introductory information similar to the sample letter included in Section 6 for the Pilot Test Program will be provided. This material shall describe the scope and general requirements of the interlaboratory test activity and the objectives of the tests. Second, detailed instructions shall accompany each sample, and these shall include a definition of any requirement unique to the sample as well as a specification of the one or more acceptable EPA or other test methods to be used in the analysis. Finally, instructions shall be included for the acquisition and recording of the test data, and for reporting general information or comments from the individual laboratories. Completion schedules shall be specified as required.

PARTICIPATING LABORATORIES

The various EPA and State laboratories which will participate in these tests are known to possess widely differing capabilities. Differences are primarily due to availability of instrumentation and analytical equipment required for measurements near detection limits. Undoubtedly, there are also differences which are attributable to personnel skills and training and to in-house administrative and technical (primarily quality control) procedures.

When major differences exist among the capabilities of participating laboratories and sample concentrations are near trace levels, analysis of the data is made difficult and confusing. For any interlaboratory study, the sample concentrations should be chosen so that they are well within the normal detection range. In this way, the capabilities of each laboratory may be properly assessed and its specific deficiencies identified.

SECTION IX

PILOT PROGRAM

GENERAL

This section contains a summary and description of a pilot interlaboratory test program whose objective is the validation and demonstration of the plan discussed in the preceding sections of this report. Subsequent to FMC's submittal of a pilot program in draft form, the Environmental Monitoring and Support Laboratory redirected the scope of the pilot test activity. The following paragraphs retain the general outline of work and mathematical methods as originally submitted. However, these have been modified in several respects, in part because of the nature of the data used for laboratory evaluation, and in part as the results of findings developed in the data analysis.

SAMPLE COMPOSITION

Analysis of trace metals was selected for the pilot test program since the atomic absorption techniques incorporate procedures which produce results having the greatest precision and accuracy among the many procedures used in water chemistry.

The selection of the number and concentrations of the trace metals samples proposed for the pilot program were based upon several parameters. These factors included recognition of quality of the select group of laboratories that will be participating in this pilot program and the need to obtain an objective differentiation between the minor differences in their analytical ability.

A total of three pairs of samples, Table 9-1, with discrete differences in concentrations represent a compromise between the maximum number of samples to be analyzed without excessive analytical time on the part of the participating laboratories and the minimum number of samples required to prove the statistical concepts.

The samples are paired in each concentration range in order to detect bias errors using Youden techniques. A relatively high spread in the concentrations within each pair of samples at or near the detection limits of the analytical procedure has been provided since it is anticipated that the analytical results will vary widely around the "true value" and this will allow the determination of bias error.

The 13 trace metals, from a total of 28 covered by standard methods, were selected with two factors considered; their potential hazard to the environment and their potential for interference in analysis.

TABLE 9-1. SAMPLE COMPOSITIONS FOR PILOT TEST PROGRAM

Metals	Trace Level Contamination (μg/liter) 1 Low	2 High	Medium Level Contamination (μg/liter) 3 Low	4 High	Normal Water Contamination (mg/liter) 5 Low	6 High
1. Aluminum (Al)	7	23	--	--	0.5	0.9
2. Arsenic (As)	20	50	75	90	--	--
3. Cadmium (Cd)	--	--	50	75	0.3	0.55
4. Chromium (Cr)	1	5	--	--	0.8	1.4
5. Copper (Cu)	1	5	12	19	--	--
6. Iron (Fe)	--	--	5	12	1.4	3
7. Mercury (H_g)	0.5	3	25	50	--	--
8. Lead (Pb)	12	28	--	--	2.4	3.2
9. Manganese (Mn)	--	--	12	19	0.1	0.2
10. Nickel (Ni)	--	--	20	50	0.33	0.70
11. Selenium (Se)	10	25	50	70	--	--
12. Zinc (Zn)	1	5	7	22	--	--
13. Cobalt (Co)	--	--	10	30	0.22	0.35

The concentration range for each set of pairs was selected as follows:

Pair 1-2:	Metal ions at or below the detection limit unless special instrumentation is available and/or extraction process is employed.
Pair 3-4:	Metal ions at or slightly above detection limits of well instrumented laboratory. Extraction may be required in some cases.
Pair 5-6:	Concentration ranges that are reported in literature as being normal levels encountered in surface water. Interferences are present in order to detect finite differences between laboratory capabilities.

In conjunction with these samples being analyzed by the participating laboratories, it is essential that either two reference laboratories analyze 6 replicates or one reference laboratory analyze 10 replicates of each of these six samples. The means and standard deviations of these data will be used as a baseline for performance evaluation.

INSTRUCTIONAL MATERIAL

Instructional material for the pilot test program is contained in the Appendix. This material includes a cover letter of general instructions, a list of consitituents to be analyzed, test methods to be used, and forms and instructions for recording and processing test data within the laboratory.

Test and reporting examples are included in this Appendix, and detailed instruction is provided for filling in the required data and commentary forms.

DATA ANALYSIS

For reasons of economy and convenience the Environmental Monitoring and Support Laboratory, responsible for the administration and management of the program, supplied FMC with the raw data results submitted by 18 EPA laboratories for "EPA Method Study 7, Trace Metals", in lieu of analytical results for the proposed series of tests. These data have been subjected to the mathematical treatment described in the pilot program plan.

Method Study Procedure

The objectives of this study and instructions to be followed by participating laboratories, issued by the National Environmental Research Center, Analytical Quality Control Laboratory, Cincinatti, Ohio, are reproduced in Appendix 1. Six sample concentrates were distributed, and they were to be analysed for Al, As, Cd, Cr, Cu, Fe, Mn, Pb, Se and Zn. The laboratories were instructed to analyze the samples only for those trace metals regularly analysed by the laboratory. Not all laboratories tested for all metals, and some tested only at certain concentrations. This procedure is a valid one when its objective is the evaluation of analytical methods. It complicates the evaluation of laboratory results when the objective is assessment of laboratory performance. Consequently, the reported results for only two metals, Cu and Zn, which 16 laboratories reported at all six concentrations, have been used in the analysis which follows.

Laboratory Data

The individual results reported by 16 laboratories are shown in Table 9-2. Data for 7 of the 10 metals to be analysed are listed in this table. Laboratory entries "Not detected" and "Not reported" are shown as a zero. Results obtained when using extraction methods are not differentiated. True values are listed for each sample group. Samples 1 and 4, 2 and 3, 5 and 6 are treated as Youden pairs.

Ordered Data and Sample Statistics

Ordered data (the "less than" and zeros are ordered among the lowest) are shown in Tables 9-3 through 9-8 in increasing rank. The corresponding laboratory numbers are shown in parentheses. Sample statistics were computed by deleting the "less than" and zeros, and these values are tabulated below the ordered data. It is seen that the sample means so computed are generally biased high compared to true values, caused by a few extremely high measurments. This also results in high standard deviations and large ranges. Relative error, defined as the difference between sample mean and true value, normalized by true value, is also shown in the parentheses.

Outlier Processing

When evaluating the analytical method, outliers would ordinarily be screened out to avoid this bias and dispersion.

TABLE 9-2. BASIC STUDY DATA FOR EPA METHOD (μg/l)

Lab No.	Cd	Cr	Cu	Fe	Pb	Mn	Zn
				Sample 1			
2	262.4	190.0	517.5	645.0	570.0	411.0	589.0
3	0.	295.0	275.0	900.0	400.0	0.	275.0
5	0.	0.	327.0	874.0	396.0	437.0	306.0
6	69.0	366.0	293.0	837.0	350.0	428.0	273.0
7	55.0	180.0	370.0	6700.0	300.0	0.	246.0
8	90.0	508.0	306.0	740.0	578.0	767.0	341.0
9	65.0	365.0	272.0	850.0	385.0	408.0	285.0
10	70.0	400.0	320.0	660.0	370.0	470.0	290.0
11	70.0	370.0	300.0	850.0	400.0	420.0	300.0
12	63.0	380.0	279.0	784.0	420.0	440.0	273.0
13	48.0	360.0	270.0	720.0	0.	340.0	260.0
14	74.0	392.0	285.0	0.	285.0	465.0	249.0
15	65.0	380.0	318.0	840.0	325.0	440.0	270.0
16	70.0	350.0	280.0	0.	400.0	400.0	260.0
17	<250.0	400.0	300.0	800.0	<500.0	430.0	270.0
18	71.0	0.	300.0	0.	387.0	394.0	282.0
True Value	71.0	370.0	302.0	840.0	367.0	426.0	281.0
				Sample 4			
2	296.5	179.0	540.5	625.0	584.0	450.0	425.0
3	0.	350.0	300.0	770.0	450.0	0.	310.0
5	0.	0.	365.0	734.0	183.0	505.0	344.0
6	77.0	392.0	326.0	697.0	320.0	469.0	300.0
7	63.0	190.0	290.0	6000.0	280.0	0.	252.0
8	77.0	568.0	350.0	610.0	520.0	873.0	359.0
9	73.0	410.0	340.0	720.0	325.0	447.0	320.0
10	70.0	360.0	300.0	850.0	400.0	450.0	280.0
11	80.0	400.0	340.0	680.0	290.0	470.0	310.0
12	72.0	420.0	303.0	651.0	355.0	490.0	293.0
13	68.0	430.0	320.0	670.0	0.	420.0	305.0
14	82.0	418.0	310.0	0.	275.0	519.0	257.0
15	72.0	400.0	345.0	700.0	370.0	480.0	310.0
16	60.0	40.0	320.0	0.	300.0	450.0	300.0
17	<250.0	400.0	300.0	700.0	<500.0	450.0	310.0
18	76.0	0.	328.0	0.	326.0	430.0	310.0
True Value	78.0	407.0	332.0	700.0	334.0	469.0	310.0

TABLE 9-2 (CONTINUED). BASIC STUDY DATA FOR EPA METHOD (μg/l)

Lab No.	Cd	Cr	Cu	Fe	Pb	Mn	Zn
				Sample 2			
2	64.0	21.0	103.2	528.0	170.0	73.0	110.0
3	0.	130.0	110.0	450.0	<400.0	0.	<150.0
5	0.	0.	136.0	399.0	68.0	84.0	83,0
6	13.0	74.0	60.0	355.0	90.0	85.0	55.0
7	0.	20.0	50.0	3300.0	50.0	0.	22.0
8	16.0	< 1.0	83.0	340.0	0.	99.0	66.0
9	17.0	80.0	54.0	390.0	122.0	82.0	70.0
10	< 10.0	10.0	10.0	0.	70.0	< 10.0	< 10.0
11	20.0	70.0	60.0	350.0	93.0	90.0	60.0
12	15.0	83.0	54.0	314.0	100.0	90.0	58.0
13	< 10.0	70.0	60.0	350.0	0.	79.0	53.0
14	24.0	75.0	80.0	0.	82.0	100.0	54.0
15	14.0	85.0	68.0	350.0	71.0	82.0	48.0
16	10.0	70.0	60.0	0.	100.0	70.0	50.0
17	<250.0	100.0	100.0	400.0	<500.0	80.0	60.0
18	14.0	0.	64.0	0.	100.0	78.0	59.0
True Value	14.0	74.0	60.0	350.0	101.0	84.0	56.0
				Sample 3			
2	79.6	30.0	126.2	594.0	140.0	72.0	126.0
3	0.	100.0	100.0	425.0	<400.0	0.	<125.0
5	0.	0.	143.0	482.0	56.0	107.0	88.0
6	14.0	94.0	73.0	426.0	70.0	107.0	68.0
7	15.0	30.0	70.0	4200.0	70.0	0.	40.0
8	14.0	115.0	87.0	400.0	0.	140.0	79.0
9	18.0	105.0	67.0	440.0	95.0	105.0	75.0
10	< 10.0	20.0	10.0	0.	50.0	< 10.0	< 10.0
11	20.0	90.0	80.0	430.0	70.0	110.0	80.0
12	19.0	104.0	68.0	406.0	88.0	115.0	72.0
13	21.0	100.0	66.0	420.0	0.	98.0	76.0
14	26.0	92.0	82.0	0.	62.0	121.0	63.0
15	17.0	90.0	82.0	430.0	82.0	106.0	75.0
16	< 10.0	100.0	70.0	0.	100.0	100.0	60.0
17	<250.0	100.0	<100.0	500.0	<500.0	90.0	70.0
18	19.0	0.	78.0	0.	83.0	94.0	75.0
True Value	18.0	93.0	75.0	438.0	84.0	106.0	70.0

TABLE 9-2 (CONTINUED). BASIC STUDY DATA FOR EPA METHOD (μg/l)

Lab No.	Cd	Cr	Cu	Fe	Pb	Mn	Zn
				Sample 5			
2	19.0	0.	20.0	50.0	62.0	0.	98.0
3	22.0	0.	<100.0	<100.0	<100.0	<400.0	<150.0
5	0.	0.	29.0	37.0	50.0	15.0	21.0
6	2.0	8.0	7.0	26.0	40.0	12.0	4.0
7	0.	0.	15.0	400.0	37.0	0.	0.
8	7.0	< 1.0	18.0	48.0	0.	27.0	30.0
9	4.2	< 30.0	7.0	18.0	33.0	9.0	23.0
10	< 10.0	70.0	50.0	360.0	140.0	80.0	50.0
11	10.0	10.0	10.0	20.0	39.0	10.0	10.0
12	2.0	11.0	9.0	25.0	35.0	10.0	11.0
13	< 10.0	8.0	13.0	38.0	0.	8.0	9.0
14	2.0	6.0	9.4	0.	38.0	3.5	13.0
15	1.0	8.0	12.0	30.0	35.0	12.0	4.0
16	0.	0.	1.2	0.	4.3	0.	0.8
17	<250.0	<100.0	<100.0	<100.0	<500.0	< 20.0	< 20.0
18	< 2.0	0.	16.0	0.	50.0	12.0	9.0
True Value	1.4	7.4	7.5	24.0	37.0	11.0	7.0
				Sample 6			
2	23.6	0.	27.8	38.0	36.0	0.	97.0
3	22.0	<100.0	<100.0	<100.0	<400.0	0.	<150.0
5	0.	0.	21.0	19.0	32.0	21.0	13.0
6	3.0	16.0	13.0	13.0	30.0	18.0	8.0
7	0.	0.	18.0	200.0	30.0	0.	0.
8	8.0	< 1.0	19.0	54.0	0.	27.0	38.0
9	1.7	< 30.0	9.5	10.0	32.0	12.0	18.0
10	< 10.0	90.0	70.0	420.0	90.0	110.0	70.0
11	0.	20.0	20.0	10.0	21.0	20.0	10.0
12	7.0	17.0	13.0	12.0	16.0	15.0	13.0
13	18.0	15.0	17.0	< 10.0	0.	16.0	11.0
14	2.8	11.0	4.4	0.	25.0	14.0	14.0
15	3.0	13.0	14.0	10.0	24.0	16.0	20.0
16	0.	0.	1.0	0.	3.0	0.	1.2
17	<250.0	<100.0	<100.0	<100.0	<500.0	< 20.0	< 20.0
18	7.0	0.	18.0	0.	26.0	16.0	10.0
True Value	2.8	15.0	12.0	10.0	25.0	17.0	11.0

TABLE 9-3. SAMPLE STATISTICS: SAMPLE 1

	Cd	Cr	Cu	Fe	Pb	Mn	Zn
	Ordered data and statistics						
	<250. (17)	0. (5)	270. (13)	0. (14)	<500. (17)	0. (3)	246. (7)
	0. (5)	0. (18)	272. (9)	0. (16)	0. (13)	0. (7)	249. (14)
	0. (3)	280. (7)	275. (3)	0. (18)	285. (14)	340. (13)	260. (13)
	48. (13)	190. (2)	279. (12)	645. (2)	300. (7)	394. (18)	260. (16)
	55. (7)	295. (3)	280. (16)	660. (10)	325. (15)	400. (16)	270. (15)
	63. (12)	350. (16)	285. (14)	720. (13)	350. (6)	408. (9)	270. (17)
	65. (15)	360. (13)	298. (6)	740. (8)	370. (10)	411. (2)	273. (12)
	65. (9)	365. (9)	300. (11)	784. (12)	385. (9)	420. (11)	273. (6)
	69. (6)	366. (6)	300. (17)	800. (17)	387. (18)	428. (6)	275. (3)
	70. (10)	370. (11)	300. (18)	837. (6)	396. (5)	430. (17)	282. (18)
	70. (16)	380. (12)	306. (8)	840. (15)	400. (16)	437. (5)	285. (9)
	70. (11)	380. (15)	318. (15)	850. (11)	400. (11)	440. (15)	290. (10)
	71. (18)	392. (14)	320. (10)	850. (9)	400. (3)	440. (12)	300. (11)
	74. (14)	400. (17)	327. (5)	874. (5)	420. (12)	465. (14)	306 (5)
	90. (8)	400. (10)	370. (7)	900. (3)	570. (2)	470. (10)	341. (8)
	262. (2)	503. (8)	513. (2)	6700. (7)	578. (8)	767. (8)	589. (2)
	Sample statistics						
Mean, M	82.49	352.57	313.28	1246.15	397.57	446.43	298.06
Relative error, RE	0.16	-0.05	0.04	0.48	0.08	0.05	0.06
Standard deviation, SD	54.94	84.10	60.16	1640.62	84.84	97.71	81.00
Range, R	214.40	328.00	247.50	6055.00	293.00	427.00	343.00
99% Ranges	4 15	3 16	1 15	4 15	3 16	3 15	1 15
	Sample statistics after deleting data beyond 99% range						
Mean, M	67.50	352.57	299.67	791.67	397.57	421.77	278.67
Relative error, RE	-0.05	-0.05	-0.01	-0.06	0.08	-0.01	-0.01
Standard deviation, SD	10.23	84.10	26.47	83.39	84.84	33.46	24.11
Range, R	42.00	328.00	100.00	255.00	293.00	130.00	95.00
Results of Dixon's outlier test	6 14	3 16	1 14	4 15	3 14	3 15	1 15
Mean, M	68.56	352.57	294.64	791.67	368.17	421.77	278.67
Relative error, RE	-0.03	-0.05	-0.02	-0.06	0.00	-0.01	-0.01
Standard Deviation, SD	3.50	84.10	18.63	83.39	43.59	33.46	24.11
Range, R	11.00	328.00	57.00	255.00	135.00	130.00	95.00

TABLE 9-4. SAMPLE STATISTICS: SAMPLE 2

	Cd	Cr	Cu	Fe	Pb	Mn	Zn
				Ordered data and statistics			
	<250. (17)	< 1. (8)	10. (10)	0. (10)	<500. (17)	< 10. (10)	<150. (3)
	< 10. (13)	0. (5)	50. (7)	0. (14)	<400. (3)	0. (7)	< 10. (10)
	< 10. (10)	0. (18)	54. (9)	0. (16)	0. (13)	0. (3)	22. (7)
	0. (3)	10. (10)	54. (12)	0. (18)	0. (8)	70. (16)	48. (15)
	0. (5)	20. (7)	60. (11)	314. (12)	50. (7)	73. (22)	50. (16)
	0. (7)	21. (2)	60. (6)	340. (8)	68. (5)	78. (18)	53. (13)
	10. (16)	70. (13)	60. (13)	350. (15)	70. (10)	79. (13)	54. (14)
	13. (6)	70. (16)	60. (26)	350. (13)	71. (15)	80. (17)	55. (6)
	14. (15)	70. (11)	64. (18)	350. (11)	82. (14)	82. (15)	58. (12)
	14. (18)	74. (6)	68. (15)	355. (6)	90. (6)	82. (9)	59. (18)
	15. (12)	75. (14)	80. (14)	390. (9)	93. (11)	84. (5)	60. (17)
	16. (8)	80. (9)	83. (8)	399. (5)	100. (16)	85. (6)	60. (11)
	17. (9)	83. (12)	100. (17)	400. (17)	100. (12)	90. (12)	66. (8)
	20. (11)	85. (15)	102. (2)	450. (3)	100. (18)	90. (11)	70. (9)
	24. (14)	100. (17)	110. (3)	528. (2)	122. (9)	99. (3)	83. (5)
	64. (2)	130. (3)	136. (5)	3300. (7)	170. (2)	100. (14)	110. (2)
				Sample statistics			
Mean, M	20.70	68.31	72.01	627.17	93.00	84.00	60.57
Relative error, RE	0.48	-0.08	0.20	0.79	-0.08	0.00	0.08
Standard deviation, SD	15.70	33.54	29.57	843.71	30.95	8.93	19.54
Range, R	54.00	120.00	126.00	2986.00	120.00	30.00	88.00
99% Ranges	7 15	4 16	1 16	5 15	5 16	4 16	3 16
				Sample statistics after deleting data beyond 99% range			
Mean, M	15.89	68.31	72.01	384.18	93.00	84.00	60.57
Relative error, RE	0.13	-0.08	0.20	0.10	0.08	0.00	-0.08
Standard Deviation, SD	4.11	33.54	29.57	60.62	30.95	8.93	19.54
Range, R	14.00	120.00	126.00	214.00	120.00	30.00	88.00
Results of Dixon's outlier test	6 15	3 16	3 14	5 14	5 16	4 16	4 15
Mean, M	18.30	83.57	76.92	435.90	80.50	105.00	73.42
Relative error, RE	0.02	-0.10	0.03	-0.00	-0.04	-0.0.	0.06
Standard deviation, SD	3.65	31.61	10.02	31.58	24.12	16.13	7.59
Range, R	12.00	95.00	34.00	100.00	90.00	68.00	28.00

TABLE 9-5. SAMPLE STATISTICS: SAMPLE 3

	Cd	Cr	Cu	Fe	Pb	Mn	Zn
	Ordered data and statistics						
	<250. (17)	0. (5)	<100. (17)	0. (10)	<500. (17)	< 10. (10)	<125. (3)
	< 10. (16)	0. (18)	10. (10)	0. (14)	<400. (3)	0. (7)	< 10. (10)
	< 10. (10)	20. (10)	66. (13)	0. (16)	0. (13)	0. (3)	40. (7)
	0. (5)	30. (2)	67. (9)	0. (18)	0. (8)	72. (2)	60. (16)
	0. (3)	30. (7)	68. (12)	400. (8)	50. (10)	90. (17)	63. (14)
	14. (8)	90. (15)	70. (16)	406. (12)	56. (5)	94. (18)	68. (6)
	14. (6)	90. (11)	70. (7)	420. (13)	62. (14)	98. (13)	70. (17)
	15. (7)	92. (14)	73. (6)	425. (3)	70. (6)	100. (16)	72. (12)
	17. (15)	94. (6)	78. (18)	426. (6)	70. (7)	105. (9)	75. (9)
	18. (9)	100. (13)	80. (11)	430. (11)	70. (11)	106. (15)	75. (15)
	19. (12)	100. (16)	82. (15)	430. (15)	82. (15)	107. (5)	75. (18)
	19. (18)	100. (17)	82. (24)	440. (9)	83. (18)	107. (6)	76. (13)
	20. (11)	100. (3)	87. (8)	482. (5)	88. (12)	110. (11)	79. (8)
	21. (13)	104. (12)	100. (3)	500. (17)	95. (9)	115. (12)	80. (11)
	26. (14)	105. (9)	126. (2)	594. (2)	100. (16)	121. (14)	88. (5)
	30. (2)	115. (8)	143. (5)	4200. (7)	140. (2)	140. (8)	126. (2)
	Sample statistics						
Mean, M	23.87	83.57	80.15	762.75	80.50	105.00	74.79
Relative error, RE	0.33	-0.10	0.07	0.74	-0.04	-0.01	0.07
Standard deviation, SD	18.80	31.61	29.55	1083.78	24.12	16.13	18.58
Range, R	65.60	95.00	133.00	3800.00	90.00	68.00	86.00
99% Ranges	6 15	3 16	2 16	5 15	5 16	4 16	3 15
	Sample statistics after deleting data beyond 99% range						
Mean, M	18.30	83.57	80.15	450.27	80.50	105.00	70.85
Relative error, RE	0.02	-0.10	0.07	0.03	-0.04	-0.01	0.01
Standard deviation, SD	3.65	31.61	29.55	56.30	24.12	16.13	11.77
Range, R	12.00	95.00	133.00	194.00	90.00	68.00	48.00
Results of Dixon's outlier test	7 15	4 16	1 16	5 14	5 15	4 16	4 15
Mean, M	15.89	68.31	72.01	369.80	86.00	84.00	59.67
Relative error, RE	0.13	-0.08	0.20	0.06	-0.15	0.00	0.07
Standard deviation, SD	4.11	33.54	29.57	39.44	20.16	8.93	9.64
Range, R	14.00	120.00	126.00	136.00	72.00	30.00	35.00

TABLE 9-6. SAMPLE STATISTICS: SAMPLE 4

	Cd	Cr	Cu	Fe	Pb	Mn	Zn
	Ordered data and statistics						
	<250.(17)	0.(5)	290.(7)	0.(14)	<500.(17)	0.(3)	252.(7)
	0.(5)	0.(18)	300.(3)	0.(16)	0.(13)	0.(7)	257.(14)
	0.(3)	40.(16)	300.(10)	0.(18)	183.(5)	420.(13)	280.(10)
	60.(16)	179.(2)	300.(17)	610.(8)	275.(14)	430.(18)	293.(12)
	63.(7)	190.(7)	303.(12)	625.(2)	280.(7)	447.(9)	300.(6)
	68.(13)	350.(3)	320.(14)	651.(12)	290.(11)	450.(2)	300.(16)
	70.(10)	360.(10)	320.(16)	670.(13)	300.(16)	450.(16)	305.(13)
	72.(12)	392.(6)	320.(13)	680.(11)	320.(6)	450.(17)	310.(3)
	72.(15)	400.(11)	326.(6)	697.(6)	325.(9)	450.(10)	310.(15)
	73.(9)	400.(17)	328.(28)	700.(17)	326.(18)	469.(6)	310.(11)
	76.(18)	400.(15)	340.(9)	700.(15)	355.(12)	470.(11)	310.(17)
	77.(6)	410.(9)	340.(11)	720.(9)	370.(15)	480.(15)	310.(18)
	77.(8)	418.(14)	345.(15)	734.(5)	400.(10)	490.(12)	320.(9)
	80.(11)	420.(12)	350.(8)	770.(3)	450.(3)	505.(5)	344.(5)
	82.(14)	430.(13)	365.(5)	850.(10)	520.(8)	519.(14)	359.(8)
	297.(2)	568.(8)	541.(2)	6000.(7)	584.(2)	873.(8)	425.(2)
	Sample statistics						
Mean, M	89.73	354.07	336.09	1108.23	355.57	493.07	311.56
Relative error, RE	0.15	-0.13	0.01	0.58	0.06	0.05	0.01
Standard deviation, SD	62.44	132.18	58.55	1471.11	104.79	112.77	40.36
Range, R	236.50	528.00	250.50	5390.00	401.00	453.00	173.00
99% Ranges	4 15	3 16	1 15	4 15	3 16	3 15	1 15
	Sample statistics after deleting data beyond 99% range						
Mean, M	72.50	354.07	322.47	700.58	355.57	463.85	304.00
Relative error, RE	-0.07	-0.13	-0.03	0.00	0.06	-0.01	-0.02
Standard deviation, SD	6.56	132.18	22.12	64.94	104.79	28.67	27.65
Range, R	22.00	528.00	75.00	240.00	401.00	99.00	107.00
Results of Dixon's outlier test	4 15	3 16	1 15	4 15	3 16	3 15	1 15
Mean, M	72.50	354.07	322.47	700.58	355.57	463.85	304.00
Relative error, RE	-0.07	-0.13	-0.03	0.00	0.06	-0.01	-0.02
Standard deviation, SD	6.56	132.18	22.12	64.94	104.79	28.67	27.65
Range, R	22.00	528.00	75.00	240.00	401.00	99.00	107.00

TABLE 9-7. SAMPLE STATISTICS: SAMPLE 5

	Cd	Cr	Cu	Fe	Pb	Mn	Zn
	Ordered data and statistics						
	<250.(17)	<100.(17)	<100.(3)	<100.(3)	<500.(17)	<400.(3)	<150.(3)
	< 10.(13)	< 30.(9)	<100.(17)	<100.(17)	<100.(3)	< 20.(17)	< 20.(17)
	< 10.(10)	< 1.(8)	1.(16)	0.(16)	0.(13)	0.(7)	0.(7)
	< 2.(18)	0.(7)	7.(6)	0.(14)	0.(8)	0.(16)	1.(16)
	0.(16)	0.(2)	7.(9)	0.(18)	4.(16)	0.(2)	4.(6)
	0.(5)	0.(3)	9.(12)	18.(9)	33.(9)	4.(14)	4.(15)
	0.(7)	0.(16)	9.(14)	20.(11)	35.(12)	8.(13)	9.(13)
	1.(15)	0.(5)	10.(11)	25.(12)	35.(15)	9.(9)	9.(18)
	2.(6)	0.(18)	12.(15)	26.(6)	37.(7)	10.(11)	10.(11)
	2.(14)	6.(14)	13.(13)	30.(15)	38.(14)	10.(12)	10.(12)
	2.(12)	8.(13)	15.(7)	37.(5)	29.(11)	12.(6)	13.(14)
	4.(9)	8.(6)	16.(18)	38.(13)	40.(6)	12.(15)	21.(5)
	7.(8)	8.(15)	18.(8)	48.(8)	50.(5)	12.(18)	23.(9)
	10.(11)	10.(11)	20.(2)	50.(2)	50.(18)	15.(5)	30.(8)
	19.(2)	11.(12)	29.(5)	360.(10)	62.(2)	27.(8)	50.(10)
	22.(3)	70.(10	50.(10)	400.(7)	142.(10)	30.(10)	98.(2)
	Sample statistics						
Mean, M	7.70	17.29	15.47	95.64	46.94	18.05	21.75
Relative error, RE	4.50	1.34	1.06	2.98	0.27	0.64	2.11
Standard deviation, SD	7.84	23.30	12.02	141.26	32.32	21.36	26.47
Range, R	20.90	64.00	48.80	382.00	135.70	76.50	97.20
99% Ranges	8 16	10 16	3 15	6 16	5 15	6 15	4 15
	Sample statistics after deleting data beyond 99% range						
Mean, M	7.70	17.29	12.82	95.64	38.48	11.85	15.40
Relative error, RE	4.50	1.34	0.71	2.98	0.04	0.08	1.20
Standatd deviation, SD	7.84	23.30	7.03	141.26	14.30	6.14	13.85
Range, R	20.90	64.00	27.80	382.00	57.70	23.50	49.20
Results of Dixon's outlier test	8 16	10 15	3 15	6 14	6 15	7 14	4 14
Mean, M	7.70	8.50	12.82	32.44	41.90	11.00	12.25
Relative error, RE	4.50	0.15	0.71	0.35	0.13	0.	0.75
Standard deviation, SD	7.84	1.76	7.03	11.56	9.19	2.20	8.96
Range, R	20.90	5.00	27.80	32.00	29.00	7.00	29.20

TABLE 9-8. SAMPLE STATISTICS: SAMPLE 6

	Cd	Cr	Cu	Fe	Pb	Mn	Zn
	<250.(17)	<100.(3)	<100.(3)	<100.(3)	<500.(17)	< 20.(17)	<150.(3)
	< 10.(10)	<100.(17)	<100.(17)	<100.(17)	<400.(3)	0.(3)	< 20.(17)
	0.(5)	< 30.(9)	1.(16)	< 10.(13)	0.(13)	0.(7)	0.(7)
	0.(11)	< 1.(8)	4.(14)	0.(16)	0.(8)	0.(16)	1.(16)
	0.(16)	0.(2)	10.(9)	0.(14)	3.(16)	0.(2)	8.(6)
	0.(7)	0.(5)	13.(12)	0.(18)	16.(12)	12.(9)	10.(11)
	2.(9)	0.(16)	13.(6)	10.(9)	21.(11	14.(14)	10.(18)
	3.(14)	0.(7)	14.(15)	10.(11)	24.(15)	15.(12)	11.(13)
	3.(6)	0.(18)	17.(13)	10.(15)	25.(14)	16.(13)	13.(5)
	3.(15)	11.(14)	18.(7)	12.(12)	26.(18)	16.(15)	13.(12)
	7.(12)	13.(15)	18.(18)	13.(6)	30.(7)	16.(18)	14.(14)
	7.(18)	15.(13)	19.(8)	19.(5)	30.(6)	18.(6)	18.(9)
	8.(8)	16.(6)	20.(11)	38.(2)	32.(5)	20.(11)	20.(15)
	18.(13)	17.(12)	21.(5)	54.(8)	32.(9)	21.(5)	38.(8)
	22.(3)	20.(11)	28.(2)	200.(7)	36.(2)	27.(8)	70.(10)
	24.(2)	90.(10)	70.(10)	420.(10)	90.(10)	110.(10)	97.(2)
	Sample statistics						
Mean, M	9.61	26.00	18.99	78.60	30.42	25.91	24.86
Relative error, RE	2.43	0.73	0.58	6.86	0.22	0.52	1.26
Standard deviation, SD	8.38	28.37	16.20	133.31	20.73	28.18	27.93
Range, R	21.90	79.00	68.90	410.00	87.00	98.00	95.80
99% Ranges	7 16	10 16	3 15	7 16	5 15	6 15	4 15
	Sample statistics after deleting data beyond 99% range						
Mean, M	9.61	26.00	15.06	78.60	25.00	17.50	18.85
Relative error, RE	2.43	0.73	0.26	6.86	0.00	0.03	0.71
Standard deviation, SD	8.38	28.37	7.12	133.31	9.23	4.28	18.40
Range, R	21.90	79.00	26.70	410.00	33.00	15.00	68.80
Results of Dixon's outlier test	7 16	10 15	3 15	7 14	6 15	6 15	4 13
Mean, M	9.61	15.33	15.06	20.75	27.20	17.50	11.82
Relative error, RE	2.43	0.02	0.26	1.08	0.09	0.03	0.07
Standard deviation, SD	8.38	3.14	7.12	16.43	5.96	4.28	5.24
Range, R	21.90	9.00	26.70	44.00	20.00	15.00	18.80

For laboratory performance evaluation any result, however far removed it may be from its nearest neighbor or the centroid of the data, must be retained to arrive at a score for each laboratory. Therefore, the raw data at this point, were not subjected to Dixon's or Grubb's procedures for outlier rejection. A less restrictive screening procedure was applied, namely rejection of values lying outside a 99 percent range centered about the sample mean. The rank of ordered results retained after this procedure is tabulated for each element. Using only the retained values, sample mean and standard deviation were recomputed, and these values are used in the Thompson ranking procedure described below.

For purposes of comparison, Dixon's outlier processing was also performed. The integral numbers again indicate the ordered positions within which the data satisfy the Dixon criterion. In other words, the data beyond the ordered positions can be treated as outliers at 95% confidence level. Statistical computations carried out after such an outlier processing are shown in the same table, which produce relative errors similar to those resulting from the previous data screening process.

It is emphasized that apparent outliers should not be rejected indiscriminately, or at the same significance level one would use in method evaluation. One should not be surprised to find large standard deviation or small (2 or 3 points) clusters of reported results widely separated from the remainder of the results. Nothing inherent in the test program precludes grossly poor performance by several participants. Therefore, these anomalies were deleted only when it was felt their inclusion would so bias the computation of mean and standard deviation as to penalize good performers in subsequent computations.

Youden Two-Sample Plot and Bias Error

Youden two-sample plots were generated by computer and plotter for three sample pairs: Samples 1 and 4, 2 and 3, 5 and 6. Results for Cu and Zn were plotted, Figures 9-1 through 9-6. The true values of the two samples in each plot are signified by intersecting lines. It is seen that most laboratories reported measurements in the first and third quadrants, namely either both positive errors or both negative errors, and the points form a significantly elongated ellipse in each plot. The measurements are biased mostly high at the low concentrations (samples 5 and 6), which indicate that most laboratories make positive errors in such cases. Furthermore, there are always one or two points found far removed from the major cluster, a result often found by other researchers.

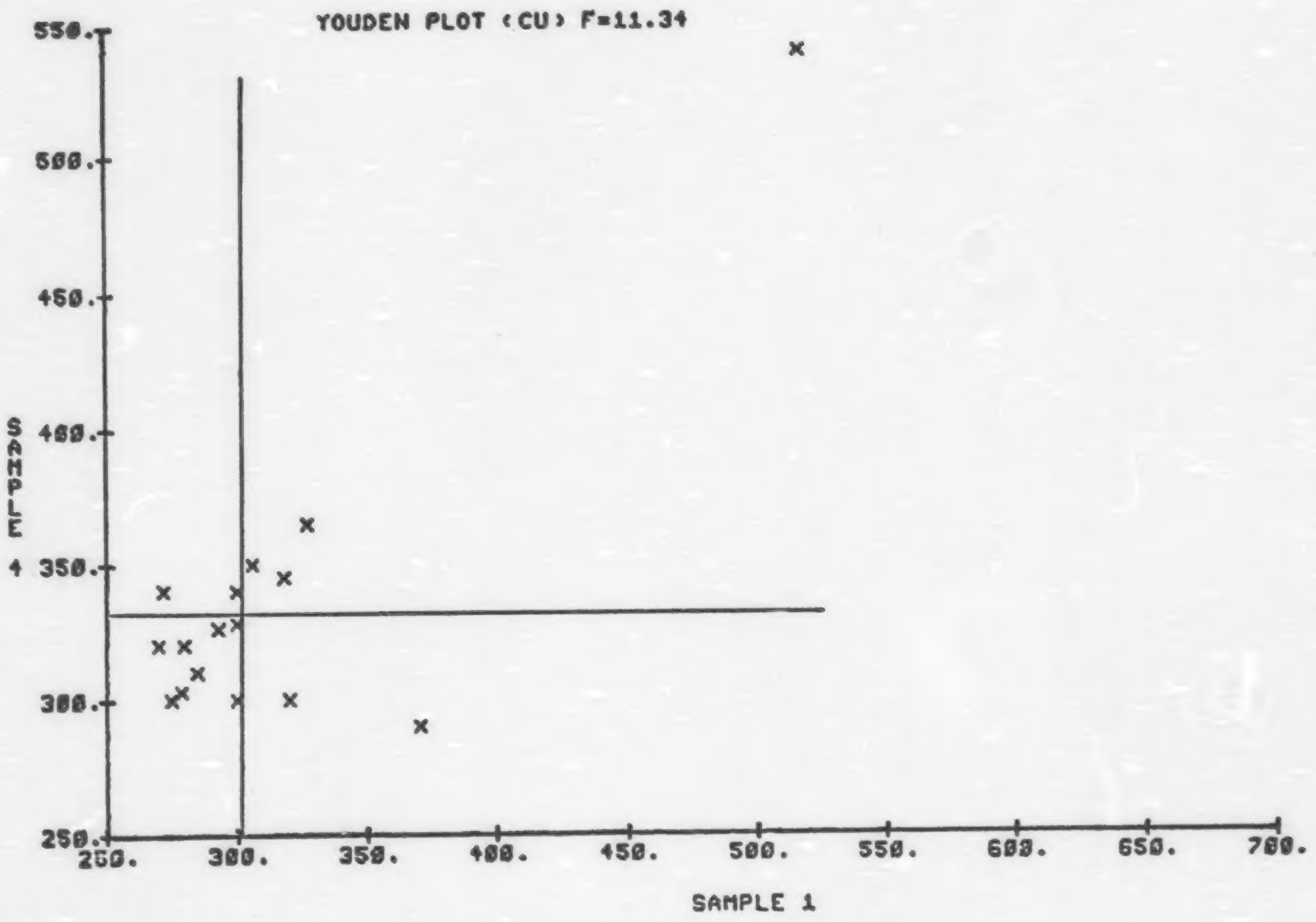

Figure 9-1. Youden's plot, Cu, samples 1 & 4.

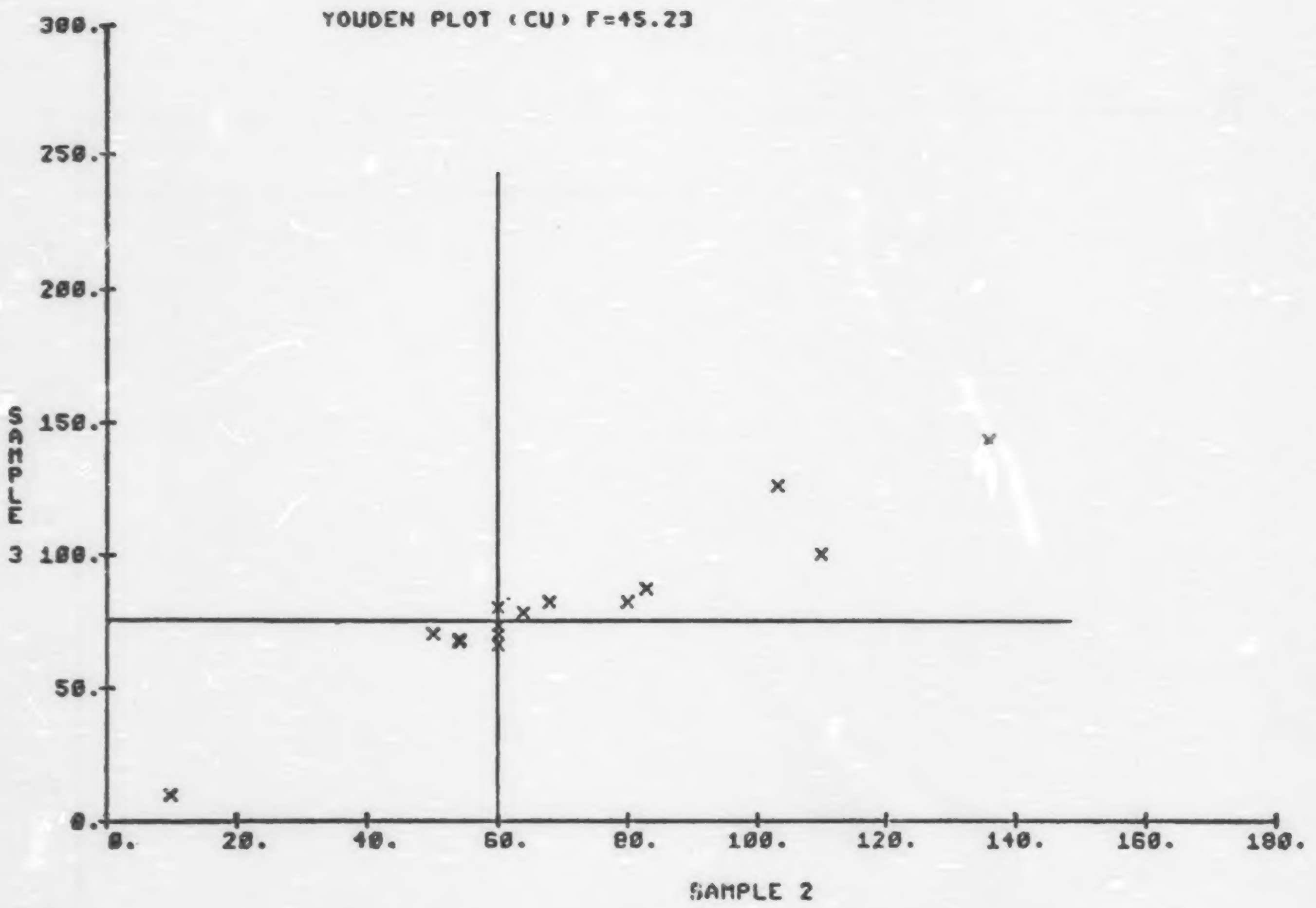

Figure 9-2. Youden's plot, Cu, samples 2 & 3.

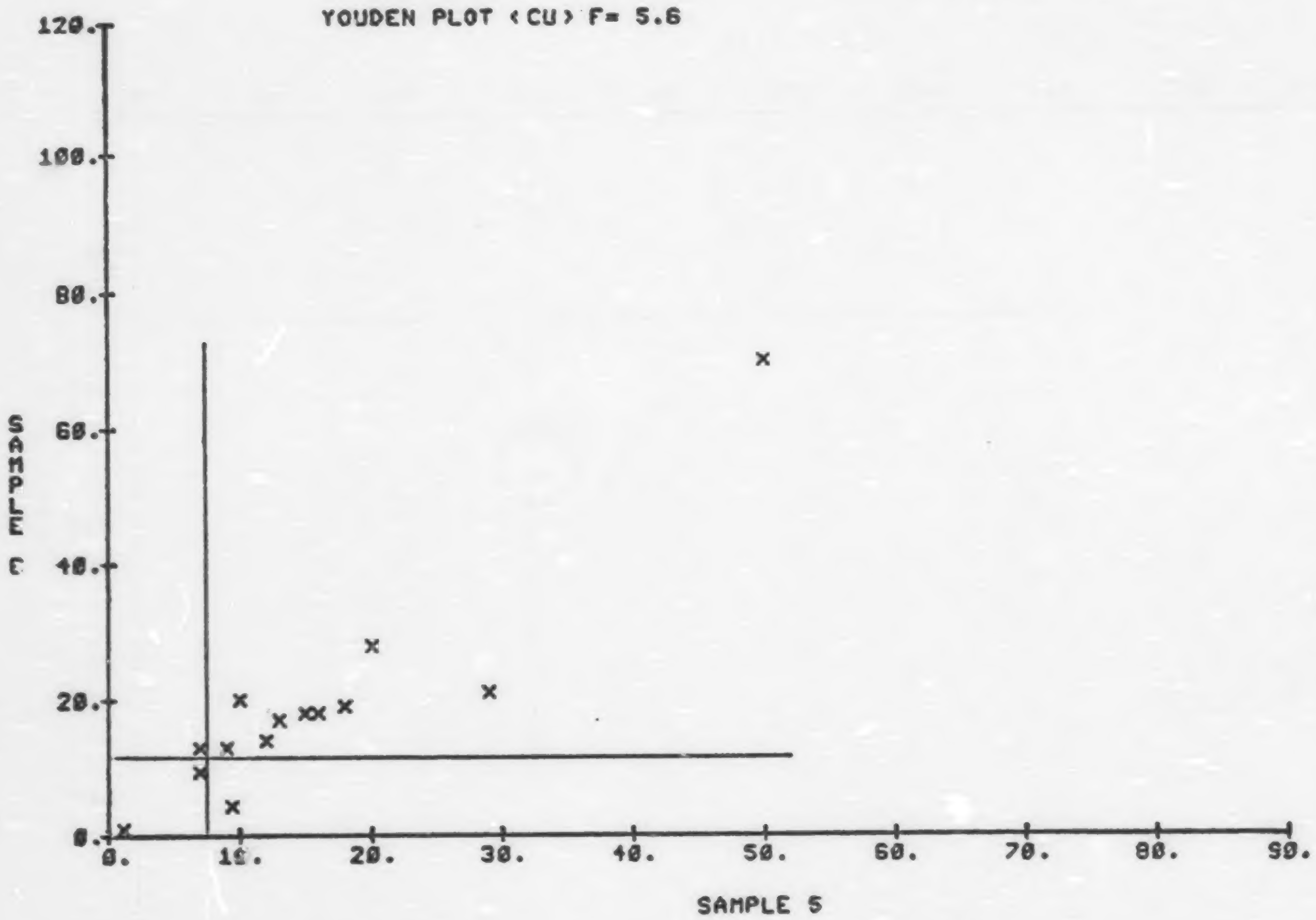

Figure 9-3. Youden's plot, Cu, samples 5 & 6.

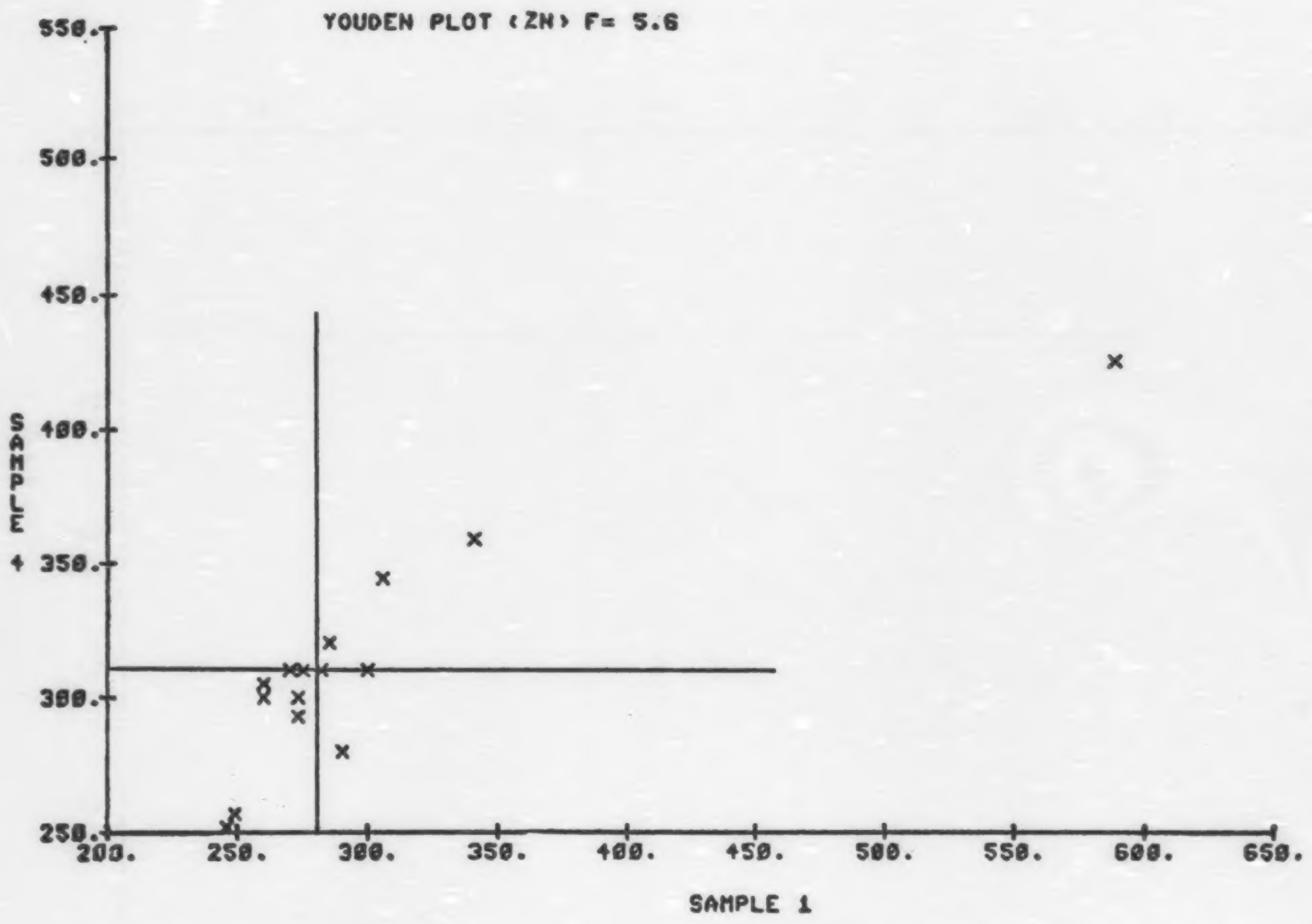

Figure 9-4. Youden's plot, Zn, samples 1 & 4.

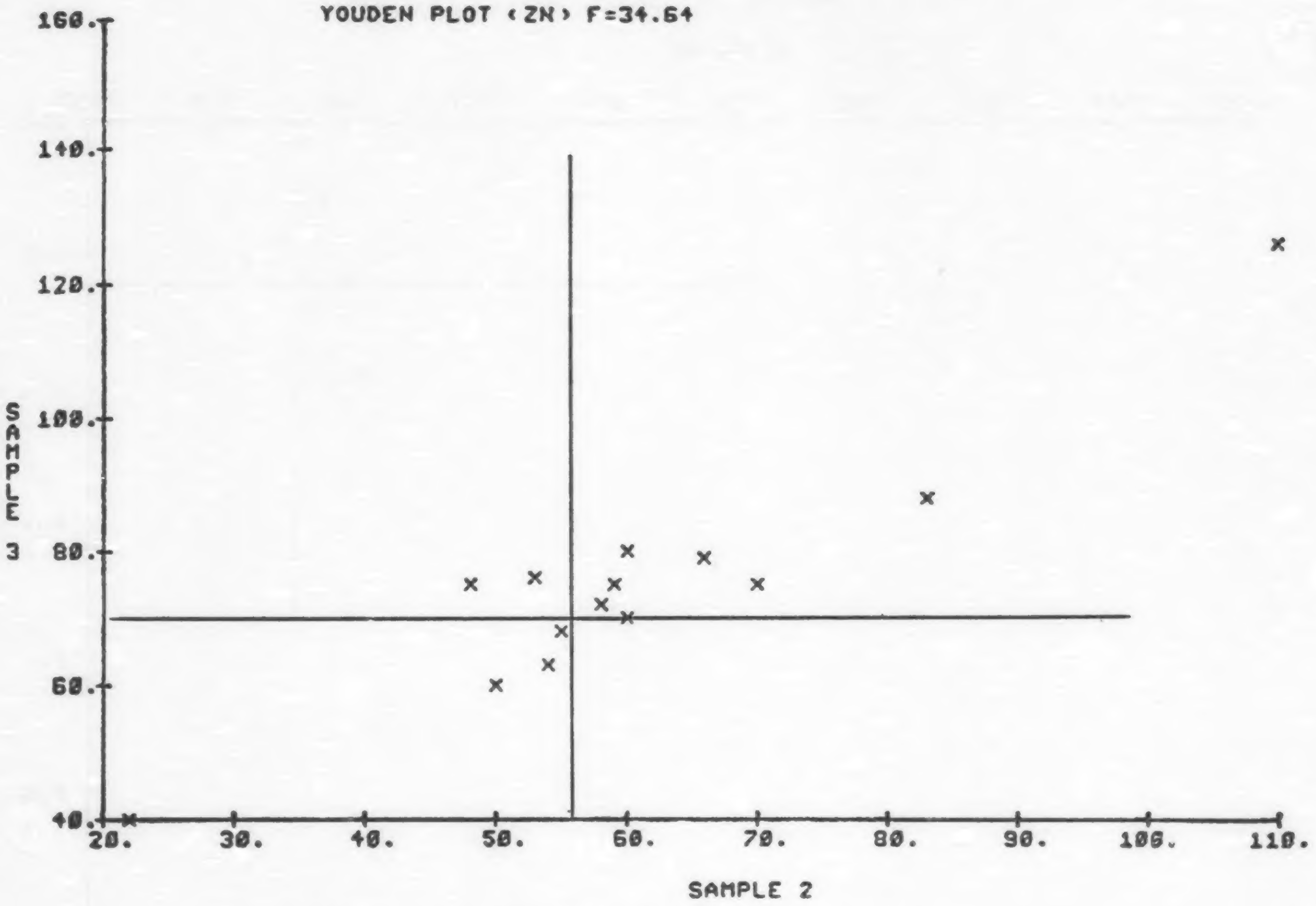

Figure 9-5. Youden's plot, Zn, samples 2 & 3.

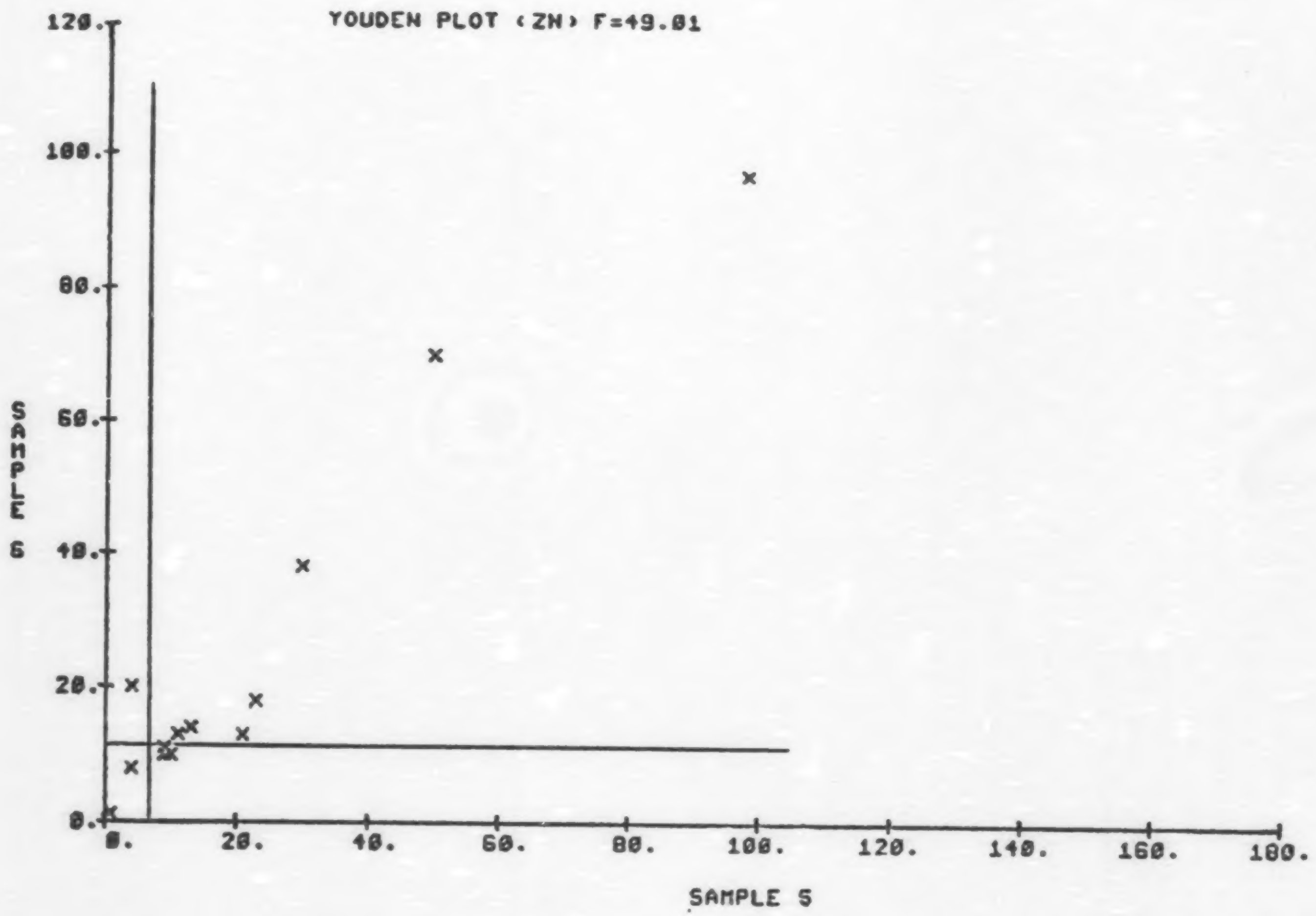

Figure 9-6. Youden's plot, Zn, samples 5 & 6.

The numerical calculation, F-ratio, suggested by Youden was also carried out, and its value is indicated along with the title of each plot, and varies from 5.6 to 49.01. Since there are 14 to 16 points plotted in each case, for which the critical F ratio at 99% confidence level lies between 3.54 and 3.93, the calculated f ratios clearly indicate bias error in each case. The critical F ratios cited above are interpolated from the critical values given by Youden, i.e., 4.16 at 12 degrees of freedom (DF), 3.70 at 14 DF, and 3.37 at 16 DF. It is concluded, therefore, for the elements Cu and Zn, the bias errors (also called systematic errors) are definitely (at 99% confidence level) present. One may further separate the standard deviations of bias errors (Sb^2) and random errors (Sr^2) by using the relationship $Sd^2 = 2Sb^2 + Sr^2$. However, because the major concern here is the detection of bias errors and to pursue the ranking of laboratory performance, no effort was made to estimate the magnitude of random errors.

Relative Errors

As a first step in ranking the laboratory performance, relative errors (R.E.) were used to discern the differences in laboratory performance. Relative error is defined as:

$$\text{R.E.} = \frac{\text{Measurement - True Value}}{\text{True Value}}$$

namely the normalized error. The R.E.'s for each laboratory at each concentration were computed for Cu and Zn. These values were averaged over the samples 1, 2, and 5, and the samples 4, 3, and 6 separately. The average R.E.'s are treated as a coordinate value to be plotted as a Youden Two-Sample Plot. Such distribution is shown in Figures 9-7 and 9-8 for Cu and Zn respectively. Similar features are found in these two plots;

(a) heavy distribution on the I and III quadrants,

(b) points forming an elongated ellipse and

(c) one or two points far removed from the center of the cluster.

Thompson's Ranking Method

This method is a quantitative measure of laboratory performance. It accounts for both errors in measurements and errors in identification. In this study, it was found that the Thompson score for measurement error can be modified to better

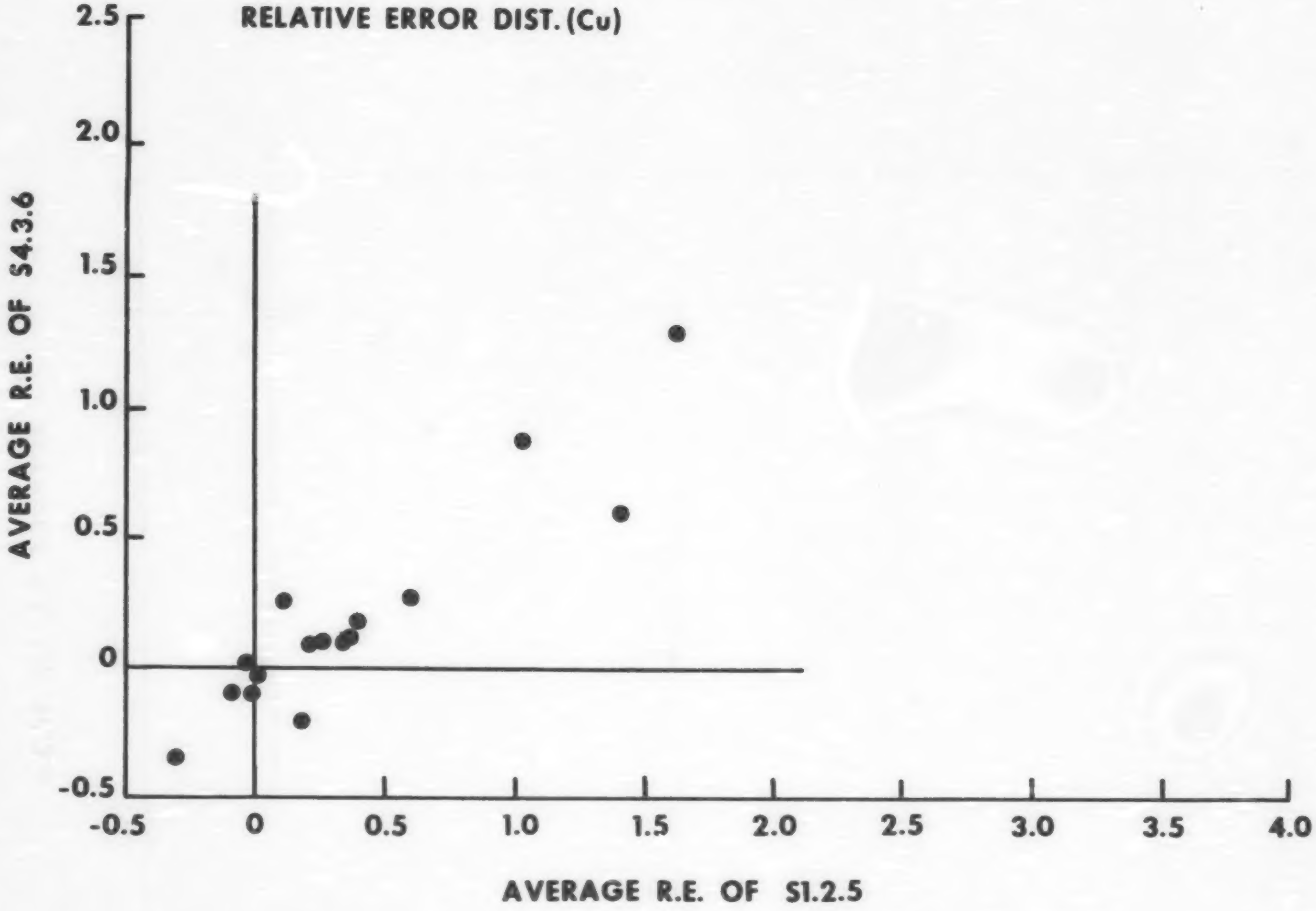

Figure 9-7. Relative errors distribution, Cu.

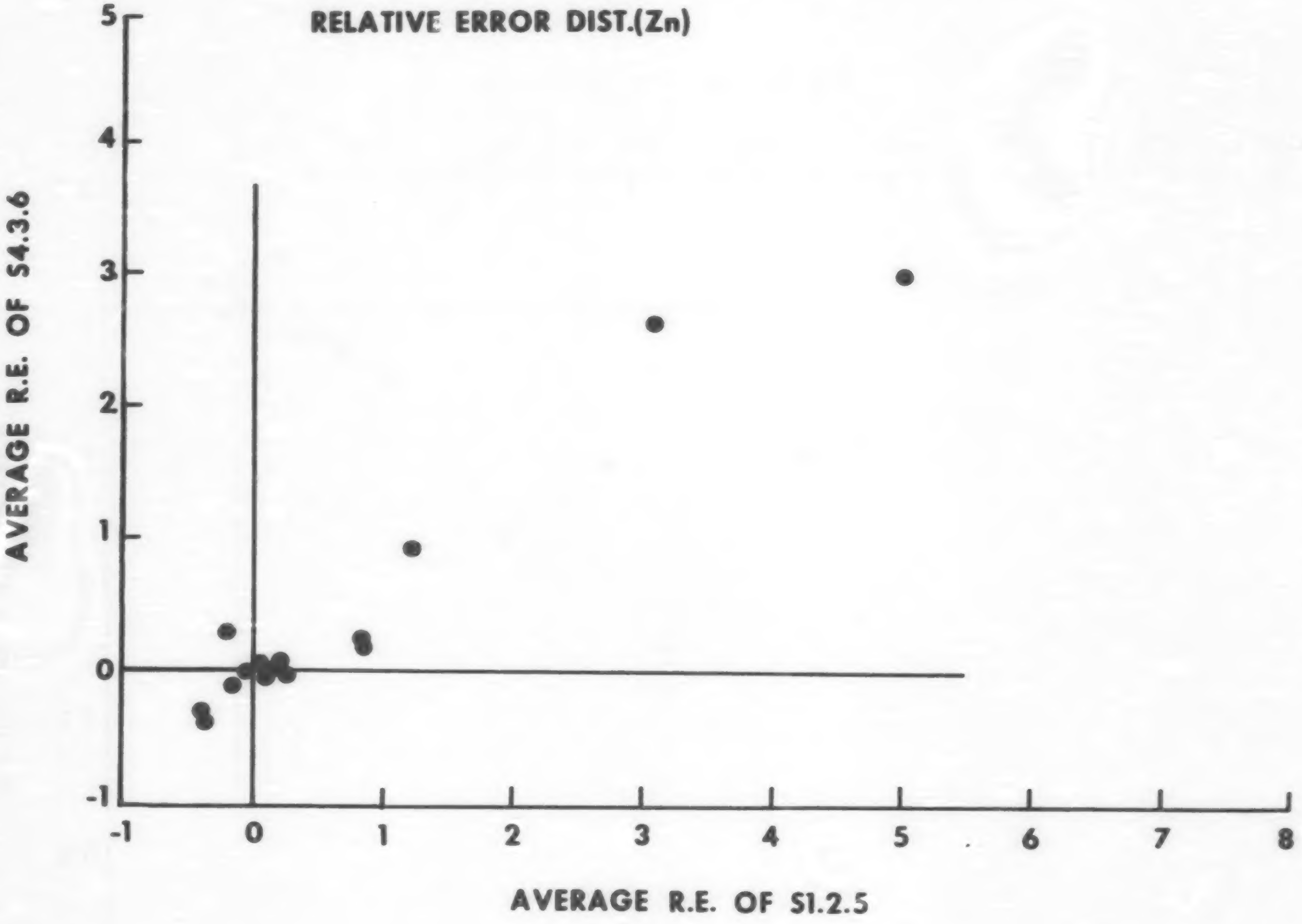

Figure 9-8. Relative errors distribution, Zn.

represent the laboratory performance. This modification is based on the observation that when the concentration level is low, the measurements tend to bias on the high side and have large sample standard deviation (SD). As the score is computed by dividing the absolute difference between the measurement (X) and the true value (TV) by the sample standard deviation, $X - TV/SD$, the magnitude of the error is reduced compared to that divided by TV, $X - TV/TV$. That is to say:

$$\frac{|X - TV|}{SD} < \frac{|X - TV|}{TV}$$

when $TV < SD$. As a result, Thompson's score under-represents the significance of the error at low concentration. On the other hand, at high concentrations, the measurements cluster more closely around TV and $SD < TV$ reversing the above inequality. In this case, Thompson's score does represent the significance of the error.

A modified Thompson's Quantification Score (PQ) is therefore used to rank the laboratory performance, which is expressed by the following equation:

$$P_Q = \sum_{i=1}^{m} \left[\frac{100}{m} - \frac{|X_i - TV|}{\min\,[SD,\ TV]} \right], m = \text{No. of measurements by each lab}$$

The denominator above is the minimum of SD and TV. The identification score (P_I) is still the same:

$$P_I = 100 - N \left(\frac{100}{m}\right),\ N = \text{No. of missed elements.}$$

Table 9-9 gives the results of such a scoring system. The full score is 100 points in each case. The total score shown in the table is the sum of the two individual scores; a full score in this case is 200 points. It is seen that laboratories 6, 9, 11, 12, 13, 14, 15, 16, and 18 all have scores above 190, whereas the rest, 2, 3, 5, 7, 8, 10, and 17 have scores below 190. The highest score is achieved by laboratory 6 and the lowest is laboratory 3.

TABLE 9-9. RESULTS OF THOMPSON'S SYSTEM

QUANTIFICATION SCORE		IDENTIFICATION SCORE		TOTAL SCORE	
2	29.96	2	100.00	2	129.96
3	44.75	3	50.00	3	94.75
5	81.01	5	100.00	5	181.01
6	97.49	6	100.00	6	197.49
7	68.61	7	83.33	7	151.94
8	84.10	8	100.00	8	184.10
9	93.02	9	100.00	9	193.02
10	50.16	10	83.33	10	133.49
11	95.55	11	100.00	11	195.55
12	95.05	12	100.00	12	195.05
13	94.46	13	100.00	13	194.46
14	91.04	14	100.00	14	191.04
15	94.84	15	100.00	15	194.84
16	91.86	16	100.00	16	191.86
17	54.80	17	58.33	17	113.13
18	96.46	18	100.00	18	196.46

A performance ranking according to the modified Thompson's Score is shown below with a dotted line separating the two groups.

LAB NO.	SCORE	RANK
6	197.5	1
18	196.5	2
11	195.6	3
12	195.1	4
15	194.8	5
13	194.5	6
9	193.0	7
16	191.9	8
14	191.0	9
8	184.1	10
5	181.0	11
7	151.9	12
10	133.5	13
2	130.0	14
17	113.1	15
3	94.8	16

Youden's Ranking Method

The Youden's performance ranking method requires that the measurements first be ordered, and scores assigned to each laboratory according to its position after the ordering. In the following computation, a score of one is assigned to the laboratory reporting the lowest value and a score of two to the laboratory reporting second lowest, and so forth. If there is a T-way tie of measurements, equal scores of S + 1/2 (T-1) are given to each of the tied laboratories, where S is the original score that would have been assigned to the lowest one. For example, if the measurements after ordering are:

15, 22, 22, 22, 24, 26, 27, 33

then the corresponding scores will be

1, 3, 3, 3, 5, 6, 7, 8

where the score 3 is computed from $2 + \frac{1}{2}(3-1) = 3$.

The results of such a ranking scheme are given in Table 9-10, where performance rankings were first computed for the elements (Cu and Zn) and the samples (1 to 6) separately. It is seen that such a breakdown generates less definitive ranking

TABLE 9-10. RESULTS OF YOUDEN'S RANKING

	By Sample						By Element		
Lab No.	1	4	2	3	5	6	Cu	Zn	Total Ranking
2	32.00	32.00	30.00	31.00	30.00	31.00	90.00	96.00	186.0
3	12.00	13.00	16.50	15.50	3.50	3.50	38.00	26.00	64.0
5	28.00	29.00	31.00	31.00	27.00	23.50	90.00	79.50	169.5
6	14.50	14.50	14.50	14.00	10.00	11.50	41.50	37.50	79.0
7	16.00	2.00	5.00	9.50	13.00	12.50	46.00	12.00	58.0
8	26.00	29.00	25.00	26.00	27.00	26.00	75.00	84.00	159.0
9	13.00	24.50	17.50	14.00	17.50	17.00	30.50	73.00	103.5
10	25.00	6.00	2.50	3.50	31.00	31.00	51.00	48.00	99.0
11	22.00	21.50	18.00	24.00	17.00	19.50	58.00	64.00	122.0
12	11.50	9.00	12.50	13.00	16.00	16.00	30.00	48.00	78.0
13	4.50	14.50	12.50	15.00	17.50	17.00	37.00	44.00	81.0
14	8.00	8.00	18.00	16.50	18.00	15.00	45.50	38.00	83.5
15	17.50	23.00	14.00	21.50	14.50	21.00	63.50	48.00	112.5
16	8.50	13.00	11.50	10.50	7.00	7.00	31.50	26.00	57.5
17	14.50	13.00	24.50	8.00	3.50	3.50	29.00	38.00	67.0
18	19.00	20.00	19.00	19.00	19.50	17.00	59.50	54.00	113.5

results. A laboratory can perform well on one element but not well on the other. However, when the separate rankings are summed to give total ranking, the results are quite similar to those by the modified Thompson's score. As seen in Table 9-6, when the laboratories are ranked by their distance from the mean score, they are again separable into two distinct groups, namely those of laboratories 6, 9, 10, 11, 12, 13, 14, 15, and 18, and 2, 3, 5, 7, 8, 16, and 17. The mean score used in above computations comes from averaging the lowest possible score, 12, and highest possible score 192 = 16 X 12, i.e., mean score =

$$12 + \frac{192 - 12}{2} = 102$$

A performance ranking based on the total score is thus given below with a dotted line separating the high and low group:

Lab No.	Separation From Mean Score	Rank
9	+1.5	1
10	-3	2
15	+10.5	3
18	+11.5	4
14	-18.5	5
11	+20	6
13	-21	7
6	-23	8
12	-24	9
....	....	....
17	-35	10
3	-38	11
7	-44	12
16	-44.5	13
18	+57	14
5	+67.5	15
2	+84	16

Although the relative ranking of laboratories within each group is not the same as by Thompson's method, since the control parameter is bias rather than accuracy, the groupings are identical except that the classification of laboratories 10 and 16 are interchanged.

LABORATORY EVALUATION

The interlaboratory test program manager is faced with two tasks in evaluating the data submitted in proficiency tests. The first test is the assessment of individual laboratory performance relative to the performance of other laboratories.

It has been shown above that two tools are available: Youden's ranking method and Thompson's ranking method. The first is a test for systematic error and the second is a test for accuracy (and for precision if more than one aliquot is provided).

The second task to be performed is the assessment of performance as a whole; are all laboratories "good", all "bad", or some "good" and some "bad". Criteria for classification should be as quantitative as is possible and conventional statistical tests, outlier tests, for example, may not necessarily apply.

Inspection of Ranking Results

The reported analytical results for copper and zinc obtained from EPA Method Study 7, Trace Metals, described in the preceding paragraphs, exhibit an interesting characteristic.

Scores obtained by the modified Thompson's ranking method were ordered and plotted in the form of a cumulative probability distribution as shown in Figure 9-9. If these points deviate from a straight line, then one suspects that they do not come from a normally distributed population. From inspection of this figure, it is concluded that there are in fact two distinct populations, one with scores in the range 94 to 184 and the other with scores in the range 190 to 198.

Hence, the test program manager is inclined to conclude that those laboratories in the lowest category performed poorly, while those in the higher category performed well. (Refer to criteria established by NERC Triangle Park, shown on the bottom of page **23)**.

Suppose, however, that all scores approximated the line with the steeper slope, with a range of scores between approximately 190 and 198. Would all of them be "acceptable", or would some lower fractile be classified "unacceptable". On the other hand, if all scores were distributed more nearly approximating the lower slope (larger standard deviation) distribution, with a range, say, between 90 and 160, then what are the acceptance criteria? The test program manager should probably conclude either that all laboratories are "bad" or that there is something wrong with the design of the test program. A potential resolution of this dilemma is shown in Figure 9-10.

In this figure, the mean score of M-n laboratories is plotted as a fraction of n, where n is the number of laboratories deleted from the calculation of the mean score. The mean score is normalized, from these data, as S/200. If the remaining mean tends to converge on $S/200 \cong 1.0$, then obviously some laboratories are "good" and some "bad". If the convergence value is appreciably smaller than 1.0, say 0.7, but convergence is evident, then

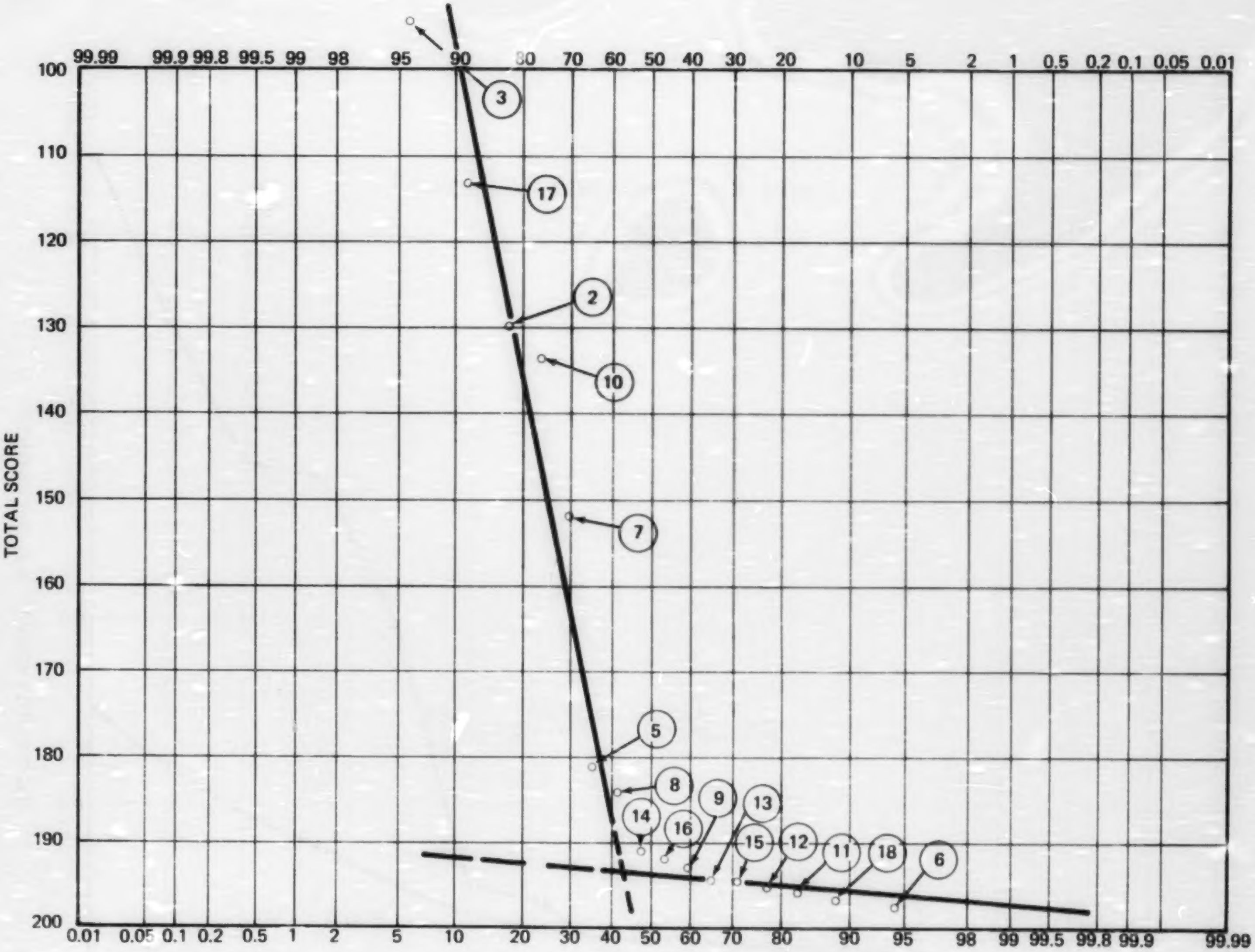

Figure 9-9. Thompson's ranking scores for 16 laboratories.

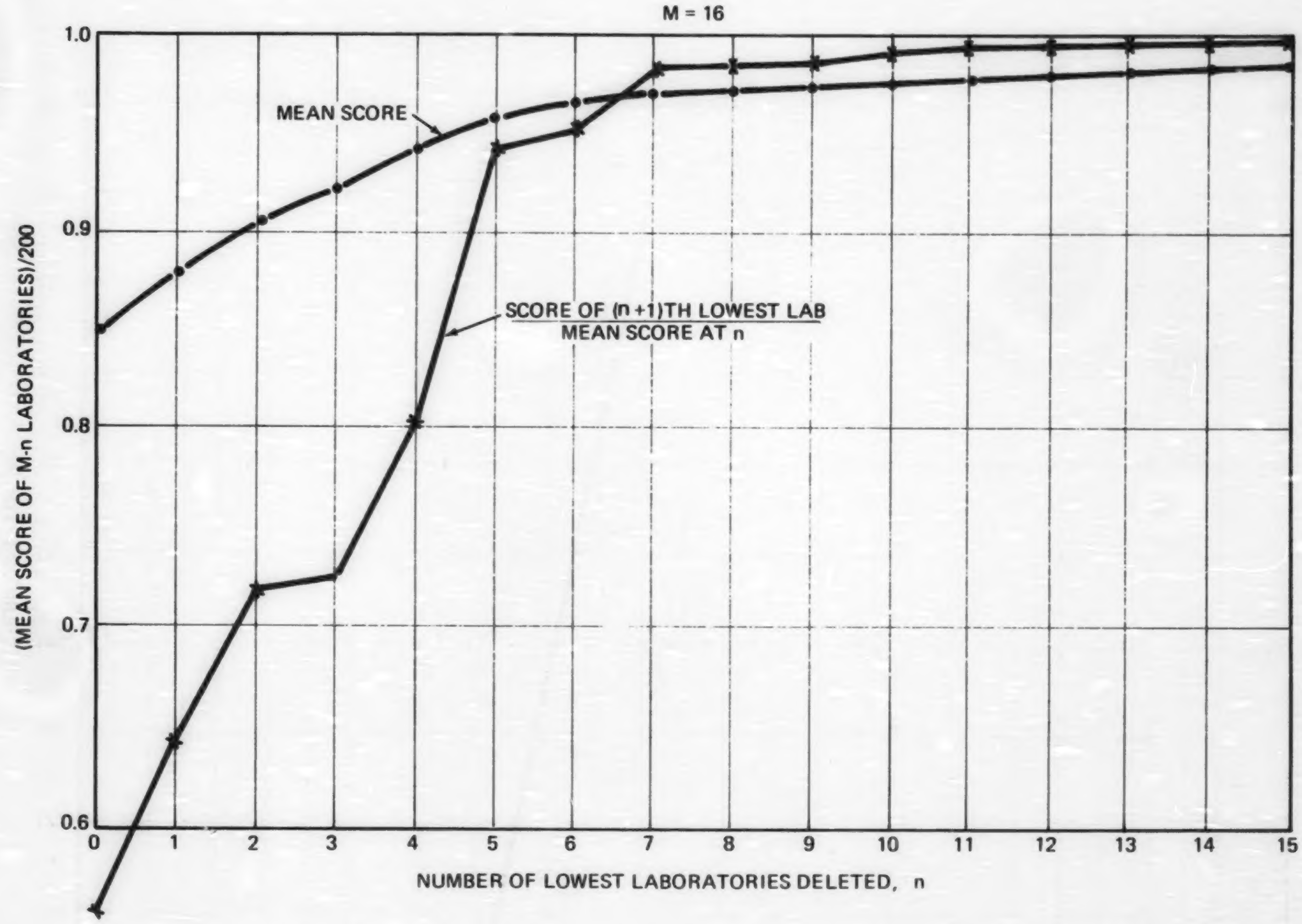

Figure 9-10. Mean score of (M-n) laboratories.

either all laboratories are "bad" or the method is "bad"; if convergence is not evident, then probably the test itself is "bad", since one does expect that regardless of circumstances some laboratories will perform better than others.

Also plotted in this figure is the ratio of the score of the (n + 1)th lowest laboratory to the mean score with n deleted. This curve illustrates the relative contribution of the (n + 1)th laboratory to degredation of the mean of the scores of the remaining laboratories. Both of the figures illustrate that laboratories 3, 17, 2, 10 and 7 are clearly unsatisfactory performers and that laboratories 5 and 8 are marginal performers.

Tests of Significance on Scores

Significance (outlier) tests may also be performed on the laboratory scores. However, the purpose of the test is not to access the probability of dispersion from the mean but rather from the nominal maximum score. Thus, the interval between the highest score and the next highest, etc., should be examined. Dixon's test was applied to the interval between laboratory 10 and laboratory 8:

$$\frac{X_2 - X_1}{X_{k-1} - X_1} = \frac{191.04 - 184.10}{196.46 - 184.10} = 0.561$$

and the critical value of 0.447 at X = 0.05 (95%) was exceeded.

Using the method of ASTM D 2777,

$$T_n = \frac{X_n - \bar{X}}{S} = \frac{184.10 - 193.3870}{3.3182} = 2.4323,$$

where $\bar{X}$ and S are computed for the top 10 scores only. This ratio exceeds the critical value of 2.29 at the 5 percent significance level. Both of these tests are applied from the top down. Values of $\bar{X}$ and S for the entire population of 16 laboratories yield meaningless statistics, since the overall distribution is clearly non-normal.

By either test, the scores of laboratories 5 and 8 appear not to belong to the higher population, and their performance should be rated unacceptable, if only two classifications are used.

These tests of significance are expected to yield results equivalent to the method outlined in "Proposal Performance Evaluation Plan", EMSL-Cincinatti, 5 December 1975. This method tests each individual laboratory results against an appropriate t or X^2 statistic, and rates acceptability of each result in terms of statistics derived from previous method studies. This approach was employed by FMC in its initial treatment of the test data, and is a preferred method when adequate statistics are available for relative standard deviation, with the sample true value end as the "mean". It was observed that in general, a laboratory which did well on one sample element usually did well on the rest, and an aggregate evaluation technique, such as the cumulative score of Thompson's method, tended both to smooth performance variations and to yield the same result regarding acceptability.

Other Observations

The data evaluated for copper and zinc exhibit another characteristic which is to be expected. Relative errors are large at low concentrations and small at high concentrations. Positive bias is generally evident, and this bias is also highest at low concentrations.

It is not an objective of this study to perform method evaluation, although such was the intent of the program from which the data were obtained. Nevertheless, it is observed that relative errors ranging from about 100 percent at low concentrations to 5 or 10 percent at higher concentrations should be expected.

It was also observed that the data were for most samples normally distributed. After screening for outliers, the data were tested by Owen's procedure, (D.B. Owen, Handbook of Statistical Tables, Addison-Wesley, 1962, p 423), which uses Kolmogoror-Smirnov one-sample statistics to test for a distribution. Basically, an empirical distribution $F_n(X)$ and an assumed continuous cumulative distribution $F(X)$ are compared, from which the highest deviation is taken to check against a critical value. For the normality test, the cumulative distribution function $F(X)$ is that of a normal distribution, also commonly called error function complement. The results of the normality tests are mostly positive, when the measurements are tested element by element and sample by sample. Out of the total 42 cases, there are only two (Fe Sample 6 and Mn Sample 4) which fail the normality test. These two cases were log-transformed and tested again, but again both fail to meet the log-normality criterion. It is concluded that the data under study are normally distributed except the two cases mentioned above.

SECTION X

REFERENCES

1. Youden, W. J., "Statistical Techniques for Collaborative Tests," Manual, Association of Official Analytical Chemists, 1969.

2. "Industrial Hygiene Laboratory Accreditation (587)," Manual, U.S. Dept. of HEW, National Institute for Occupational Safety and Health, 1974.

3. "Water-Oxygen Demand Number 2, Study No. 21," HEW, Public Health Service Publication No. 999-WP-26, 1965.

4. "Water Physics No. 1, Report No. 39," EPA, 1971.

5. "Water Trace Elements No. 2, Study No. 26," HEW, 1966.

6. "Water Chlorine (Residual) No. 2, Report No. 40," EPA, 1971.

7. "Water Nutrients No. 2, Study No. 36," HEW, Public Health Service Publication No. 2019, 1970.

8. "Water Fluoride No. 3, Study No. 33," HEW, Public Health Service Publication No. 1895, 1969.

9. "Industrial Hygiene Laboratory Accreditation," Division of Training, National Institute for Occupational Safety and Health, HEW, May 1974.

10. "Criteria for Accreditation of Industrial Hygiene Laboratories," American Industrial Hygiene Association Brochure.

11. "Evaluation of Round 30 (74-4) Results," Memo, J. H. Cavender, Chemical Reference Lab., Public Health Service, HEW, Oct. 21, 1974.

12. "Statistical Protocol for Analysis of Data from PAT Samples," W. E. Crouse, DLCD, NIOSH, HEW, July 15, 1974.

13. Statistical Protocol for the Analysis of the PAT Data," W. E. Crouse, DDCD, NIOSH, Public Health Service, HEW, August 6, 1974.

14. Pierson, R. H. and Fay, E. A., "Guidelines for Interlaboratory Testing Programs," Report for Analytical Chemists, Presented at the 135th National Meeting of The American Chemical Society at Boston.

15. Greenberg, A. E., "Use of Reference Samples in Evaluating Water Laboratories," Public Health Reports, Vol. 76, No. 9, September 1961, pp. 783-787.

16. Mandel, J. & Stiehler, R. D., "Sensitivity - A Criterion for the Comparison of Methods of Tests," National Bureau of Standards, Vol. 53, No. 3, Sept. 1954, pp. 155-159.

17. Youden, W. J., "Graphical Diagnosis of Interlaboratory Test Results," Industrial Quality Control, Vol. 15, July - June, 1958-1959, pp. 24-28.

18. Mandel, J. and Linnig, F. J., "Study of Accuracy in Chemical Analysis Using Linear Calibration Curves," Analytical Chemistry, Vol. 29, No. 5, May 1957, pp. 743-749.

19. Linnig, F. J. & Mandel, J., "Which Measure of Precision?" Analytical Chemistry, Vol. 36, No. 13, Dec. 1964, pp. 25-32.

20. "Statistical Method - Evaluation and Quality Control for the Laboratory," Training Course Manual in Computational Analysis, U.S. Department of HEW, Environmental Health Facilities, 1968.

21. Kramer, H. P. & Kroner, R. C., "Cooperative Studies on Laboratory Methodology," Journal AWWA, May 1959, pp. 607-613.

22. Greenberg, A. E., Thomas, J. S., Lee, T. W., and Gaffey, W. R., "Interlaboratory Comparisons in Water Bacteriology," Journal American Water Works Association, Vol. 59, No. 2, Feb. 1967, pp. 237-244.

23. Devine, R. F. and Partington, G. L., "Interference of Sulfate Ion on SPADNS Colorimetric Determination of Fluoride in Wastewaters," Envir. Science and Tech., Vol. 9, No. 7, July 1975, pp. 678-679.

24. McFarren, E. F., et.al., "Criterion for Judging Acceptability of Analytical Methods," Analytical Chemistry, Vol. 42, No. 3, Mar. 1970, pp. 358-365.

25. "Development of a System for Conducting Interlaboratory Tests for Water Quality and Effluent Measurements," Survey Questionary, FMC.

26. Wernimont, G., "The Design and Interpretation of Interlaborabory Test Programs," ASTM Bulletin, May 1950, pp. 45-58.

27. Greenberg, A. E., "Water Laboratory Approval Program," Presented before the Laboratory Section, American Public Health Association, Nov. 1, 1960.

28. Lee, T. G., "Interlaboratory Evaluation of Smoke Density Chamber," NBS.

29. McKee, H. C., Childers, R. E., "Collaborative Study of Reference Method for the Continuous Measurement of Carbon Monoxide in the Atmosphere," Southwest Research Institute, Houston, Texas.

30. Bingham, C. D., Whichard, J., "Evaluation of an Interlaboratory Comparison Involving Pyrocarbon and Silicon Carbide-coated Uranium-Thorium Carbide Beads," USAEC New Brunswick Laboratory, N.J.

31. Lee T. G. and Huggett C., "Interlaboratory Evaluation of the Tunnel Test (ASTME 84) Applied to Floor Coverings," Inst. for Applied Technology, NBS.

32. Weiss, C. M., and Helms, R. W., "The Interlaboratory Precision Test, An Eight Laboratory Evaluation of the Provisional Algal Assay Procedure Bottle Test," Chapel Hill Dept. of Envir. Sciences and Engineering, North Carolina Univ.

33. Merkle, E. J., et al, "Interlaboratory Comparison of Chemical Analysis of Uranium Mononitride," Lewis Res. Center, NASA.

34. "Cooperative Evaluation of Techniques for Measuring Hydrocarbons in Diesel Exhaust," Coordinating Research Council, Inc., N.Y.

35. "Cooperative Evaluation of Techniques for Measuring Hydrocarbon in Diesel Exhaust, Phase III," Coordinating Research Council, Inc., N.Y.

36. McKee, H. C., et al, "Collaborative Study of Reference Method for Determination of Sulfur Dioxide in the Atmosphere (Pararosaniline Method)," Southwest Research Inst., Houston, Texas.

37. Dixon, W. J. and Massey, F. J., Jr., Introduction to Statistical Analysis, (New York: McGraw-Hill, 1957).

38. Brownlee, K. A., Statistical Theory and Methodology in Science and Engineering,(New York: Wiley, 1960).

39. Cochran W. C. and Snedecor, G. W., Statistical Methods, (Ames, Iowa: Iowa State University Press, 1967).

40. Siegel, S., Nonparametric Statistics for the Behavioral Sciences, (New York: McGraw-Hill, 1956).

41. Mace, A. E., Sample Size Determination, R. E. Kriege Co., 1974, pp. 35-37 and 56-57.

42. David, H. A., Order Statistics, Wiley & Son Inc., 1970, p. 193.

43. Bennett, C. A. and Franklin, N. L., Statistical Analysis of Chemistry and the Chemical Industry, Wiley and Son Inc., 1954, Ch. 8.

44. Barnett, R. N. and Pinto, C. L., "Evaluation of a System for Precision Control in the Clinical Laboratory," The American Journal of Clinical Pathology, Vol. 48, No. 2, 1967, pp. 243-247.

45. Copeland, B. E. "Standard Deviation," American Journal Clinical Pathology, Vol. 27, 1957, pp. 551-557.

46. Greenberg, A. E., et al, "Chemical Reference Samples in Water Laboratories," Journal American Water Works Association, Vol. 61, No. 11, No. 1969, pp. 599-602.

47. Griffin, D. F., "Systems Control by Cumulative Sum Method," American Journal of Medical Technology, Vol. 34, No. 11, Nov. 1968, pp. 644-650.

48. Sokal, R. R., and Rohlf, J. F., "Biometry; The Principles and Practice of Statistics in Biological Research," Freeman and Comp., San Francisco, Calif.

49. Wernimont, G., "Design and Interpretation of Interlaboratory Studies of Test Methods," Analytical Chemistry, Vol. 23, No. 11, Nov. 1951, pp. 1572-1976.

50. Willits, C. O., "Standardization of Microchemical Methods and Apparatus," Analytical Chemistry, Vol. 23, No. 11, Nov. 1951, pp. 1565, 1567.

51. McArthur, D. S., et al, "Evaluation of Test Procedures," Analytical Chemistry, Vol. 26, No. 6, June 1954, pp. 1012-1018.

52. Czech, F. P., "Simplex Optimized Acetylacetone Method for Formaldehyde," Journal of AOAC, Vol. 56, No. 6, 1973, pp. 1496-1502.

53. Czech, F. P., "Simplex Optimized J-Acid Method for the Determination of Formaldehyde," Journal of AOAC, Vol. 56, No. 6, 1973, pp. 1489-95.

54. Byram, K. V. and Krawczyk, D. F., "Management System for an Analytical Chemical Laboratory," Americal Laboratory, Vol. 5., No. 1, Jan. 1973, pp. 55-62.

55. Bryam, K. V. and Krawczyk, D. F., "The Use of a Management System in Operating an Analytical Chemical Laboratory," Working Paper, EPA, Pacific Northwest Water Laboratory, Corvallis, Oregon.

56. Table of Contents, 1973 Book of ASTM Standards, Parts 23 and 30.

57. Mandel J. and Paule, R. C., "Analysis of Interlaboratory Measurements on the Vapor Pressure of Gold," Inst. for Material Research, NBS.

58. Ku, H. H., "Precision Measurement and Calibration," NBS.

59. "Correlation of Full-Flow Light Extinction Type Diesel Smokemeters by a Series of Neutral Density Filters," Coordinating Research Council, Inc., N. Y.

60. "Instrumental Analysis of Chemical Pollutants, Training Manual," Office of Water Programs, EPA.

61. "Methods for Organic Pesticides in Water and Wastewater," Analytical Quality Control Laboratory, NERC, Cinc., Ohio.

62. "Evaluation of Monitoring Methods and Instrumentation for Hydrocarbon and Carbon Monoxide in Stationary Source Emissions," Walden Research Corp., Camb., Mass.

63. Bohl, D. R., Sellero, D. E., "Statistical Evaluation of Selected Analytical Procedures," Mound Laboratory, Miamisburg, Ohio.

64. Mandel, J. and Paul, R. C., "Standard Reference Material: Analysis of Interlaboratory Measurements on the Vapor Pressures of Cadmium and Silver. (Certification of Standard Reference Materials 746 and 748)," Inst. of Material Research (401 - 937), NBS.

65. McFarren, E. F., et al, "Water Metals No. 4, Study Number 30. Report of a Study Conducted by Analytical Reference Service," Bureau of Disease Prevention and Environmental Control, PHS, Cinc., Ohio.

66. Lishka, R. J., Parker, J. H., "Water Surfactant No. 3, Study Number 32. Report of a Study Conducted by Analytical Reference Service," Bureau of Disease Prevention and Environmental Control, PHS, Cinc., Ohio.

67. Arnett, E. M., "A Chemical Information Center Experimental Station," Pittsburgh Chemical Information Center, Penn.

68. "Proceedings, Joint Conference on Prevention and Control of Oil Spills," American Petroleum Institute, N.Y.

69. Ekedahl, G., et al, "Interlaboratory Study of Methods for Chemical Analysis of Water," Journal WPCF, Vol. 47, No. 4, April 1975, pp. 858-866.

70. "Industrial Hygiene Service Laboratory Quality Control," Manual, Technical Report No. 78, U.S. Dept. of HEW, National Institute for Occupational Safety and Health, updated.

71. Lark, P. D., "Application of Statistical Analysis to Analytical Data," Analytical Chemistry, Vol. 26, No. 11, Nov. 1954, pp. 1712-1715.

72. "Quality Control in the Industrial Hygiene Laboratory," Manual, U.S. Dept. of HEW, National Institute for Occupational Safety and Health, 1971.

73. Frazier, R. P., et al, "Establish a Quality Control Program for a State Environmental Lab.," Water & Sewage Works, May 1974, pp. 54-75.

74. Frazier, R. P., et al, "Establishing a Quality Control Program for a State Environmental Laboratory," Water & Sewage Works, May 1974, pp. 54-57, 75.

75. Hoffmann, R. G. and Waid, M. F., "The Number Plus Method of Quality Control of Laboratory Accuracy," The American Journal of Clinical Pathology, Vol. 40, No. 3, Sept. 1963.

76. Barnett, R. N. and Weinberg, M. S. "Absence of Analytical Bias in a Quality Control Program," The American Journal of Clinical Pathology, Vol. 38, No. 5, Nov. 1962, pp. 468-472.

77. Hoffmann, R. G. and Waid, M. E., "The Quality Control to Laboratory Precision," American Journal of Clinical Pathology, 25: 585-594, 1955.

78. Jennings, E. R. and Levey, S., "The Use of Control Charts in the Clinical Laboratory," American Journal of Clinical Pathology, 20: 1059-1066, 1950.

79. Nelson, A. C. Jr. and Smith, F., "Guidelines for Development of a Quality Assurance Program. Reference Method for Measurement of Photochemical Oxidents," Research Triangle Inst., Durham, N.C.

80. "Guideline for Development of A Quality Assurance Program. Reference Method for the Continuous Measurement of Carbon Monoxide in the Atmosphere," Research Triangle Inst., Research Triangle Park, N. C.

81. Covell, D. F., "Computer-coupled Quality Control Procedure for Gamma-ray Scintillation Spectrometry," Naval Radiological Defense Lab., San Francisco, Calif.

82. Harley, J. H. and Volchok, H. L., "Quality Control in Radiochemical Analysis," Vsaec Health and Safety Laboratory, N. Y.

83. Ballinger, D. G., et al, "Handbook for Analytical Quality Control in Water and Wastewater Laboratory," NERC, Cinc., Ohio.

84. Robert, S., "Laboratory Quality Control Manual," Kerr Water Research Center, Ada., Oklahoma.

85. "Review of Current Literature on Analytical Methodology and Quality Control, Number 22," Analytical Methodology Information Center, Battelle Columbus Laboratory, Ohio.

86. "FWPCA Method Study 1. Mineral and Physical Analyses," Analytical Quality Control Laboratory, Federal Water Pollution Control Administration, Cinc., Ohio.

87. Meiggs, T. O., "Workshop on Sample Preparation Techniques for Organic Pollutant Analysis Held at Denver, Colorado, Oct. 2-4, 1973," National Field Investigation Center, Denver, Colo.

88. Smith, F., et at, "Guideline for Development of a Quality Assurance Program, Vol 1. Determination of Stack Gas Velocity and Volumetric Flow Rate C Type-S Pitot Tube," Research Triange Inst., Durham, N. C.

89. Bailey, L. V., Arnett, L. M., "ASP - Analysis of Synthetics Program for Quality Control Data," Savannah River Laboratory, DuPont de Nemours (E. I.) and Comp., Aiken, So. Carolina.

90. "Operational Hydromet Data Management System, Design Characteristics," North American Rockwell Information Systems Comp., Anaheim, Calif.

91. "Design and Operation of An Information Center of Analytical Methodology," Battelle Memorial Institute, Columbus, Ohio.

92. "Storage and Retrieval of Water Quality Data, Training Manual," EPA, Washington D. C.

93. Lewinger, K. L. "Studies in the Analysis of Metropolitan Water Resource Systems, Vol V: A Method of Data Reduction for Water Resources Information Storage and Retrieval," Water Resources and Marine Sciences Center, Cornell University, Ithaca, N. Y.

94. Proceedings of Conference on "Toward a Statewide Ground Water Quality Information System" and "Report of Ground Water Quality Subcommittee, Citizens Advisory Committee, Governors Environmental Quality Control," Water Resources Research Center, University of Minnesota, Minneapolis, Minn.

95. Reynolds, H. D., "An Information System for the Management of Lake Ontario," Cornell University, Ithaca, N. Y.

96. "Transport and the Biological Effects on Molybdenum in the Environment," Colorado State University, Fort Collins, Colo.

97. Ward, R. C., "Data Acquisition Systems in Water Quality Management," Colorado State University, Fort Collins, Colo.

98. Steel, T. D., "The Syslab System for Data Analysis of Historical Water - Quality Records (Basic Program)," Geological Survey, Washington D. C.

99. Lehmann, E. J., "Automatic Acquisition of Water Quality Data," National Technical Information Service, Springfield, Virg.

100. Bulkley, J. W. and Yaffee, S. L., "Factors Affecting Innovation in Water Quality Management: Implementation of the 1968 Michigan Clean Water Bond Issue," Dept. of Civil Engineering, University of Michigan, Ann Arbor, Mich.

101. Guenther, G., et al, "Michigan Water Resources Enforcement and Information System," Water Resources Commission, Dept. of Natural Resources, Lansing, Michigan.

102. "A National Overview of Existing Coastal Water Quality Monitoring," Interstate Electronics Corp., Anaheim, Calif.

103. Barrow, D. R., "SIDES: Storet Input Data Editing System," Surveillance and Analysis Division, EPA, Athens, Georgia.

104. Ho, C. Y., "Theomophysical and Electronic Properties Information Analysis Center (TEPIAC): A Continuing Systematic Program on Tables of Therphysical and Electronic Properties of Materials," Theomophysical Properties Research Center, Purdue University, Lafayette, Indiana.

105. Dubois, D. P. "STORET II:" Storage and Retrieval of Data for Open Water and Land Areas," Div. of Pollution Surveillance, Fed. Water Pollution Control Adm., Washington D. C.

106. Conley, W. and Tipton, A. R., "Part I, A Conceptual Model for a Terrestrial Ecosystem Perturbed with Sewage Effluent, With Special Reference to the Michigan State University Water Quality Management Project. Part II, A Personalized Bibliographic Retrieval Package for Resource Scientists," Dept. of Fisheries and Wildlife, Michigan State University, East Lansing, Mich.

107. Stevens, S. S., Handbook of Experimental Psychology, John Wiley, 1951, pp. 35 and 1297.

108. D. Meister, "The Problem of Human-Initiated Failures," Proced. 8th National Sym. on Rel and Q.C. pp. 234-239, Jan. 9, 1964.

109. D. Meister, "Methods of Predicting Human Reliability in Man-Machine Systems," Human Factors, 6 (6), 1964.

SECTION XI

INTERLABORATORY TEST PROGRAMS

BIBLIOGRAPHY

A. INTERLABORATORY TESTS

ANALYSIS OF INTERLABORATORY MEASUREMENTS ON THE VAPOR PRESSURE OF GOLD

National Bureau of Standards, Washington, D.C. Institute for Materials Research

AUTHOR: Paule, Robert C.; Mandel, John

ABSTRACT: A detailed statistical analysis has been made of results obtained from a series of interlaboratory measurements on the vapor pressure of gold. The Gold Standard Reference Material 745 which was used for the measurements has been certified over the pressure range 10 to the -8th to 10 to the 3rd atm. The temperature range corresponding to these pressures is 1300-2100 K. The gold heat of sublimation at 298 K and the associated standard error were found to be 87,720 ± 210 cal/mol (367,040 ± 900 J/mol). Estimates of uncertainty have been calculated for the certified temperature-pressure values as well as for the uncertainties expected from a typical single laboratory's measurements. A statistical analysis has also been made for both the second and third law methods, and for within and between laboratory components of error. Several notable differences in second and third law errors are observed.

PRECISION MEASUREMENT AND CALIBRATION. SELECTED NBS PAPERS ON STATISTICAL CONCEPTS AND PROCEDURES

National Bureau of Standards, Washington, D.C.

AUTHOR: Ku, Harry H.

ABSTRACT: This volume is one of an extended series which brings together the previously published papers, monographs, abstracts, and bibliographies by NBS authors dealing with the precision measurement of specific physical quantities and the calibration of the related metrology equipment. It deals with methodology in the generation, analysis, and interpretation of precision measurement data. It contains 40 reprints assembled in 6 sections: (1) the measurement process; (2) design of experiments in calibration; (3) interlaboratory tests; (4) functional relationships; (5) statistical treatment of measurement data; (6) miscellaneous. Each section is introduced by an interpretive foreword, and the whole is supplemented by abstracts and selected references.

INTERLABORATORY EVALUATION OF SMOKE DENSITY CHAMBER

National Bureau of Standards, Washington, D.C. Building Research Division

AUTHOR: Lee, T. G.

ABSTRACT: Results are reported of an interlaboratory (round-robin) evaluation of the smoke density chamber method for measuring the smoke generated by solid materials in fire. A statistical analysis of the results from 10 material-condition combinations and 18 laboratories is presented. For the materials tested, the median coefficient of variation of reproducibility was 7.2% under non-flaming exposure conditions and 13% under flaming exposure conditions. A discussion of errors and recommendations for improved procedures based on user experience is given. A tentative test method description is included as an appendix.

COLLABORATIVE STUDY OF REFERENCE METHOD FOR THE CONTINUOUS MEASUREMENT OF CARBON MONOXIDE IN THE ATMOSPHERE (NON-DISPERSIVE INFRARED SPECTROMETRY)

Southwest Research Institute, Houston, Texas

AUTHOR: McKee, Herbert C.; Childers, Ralph E.

ABSTRACT: Information obtained in the evaluation and collaborative testing of a reference method for measuring the carbon monoxide content of the atmosphere is presented. The method is based on the infrared absorption characteristics of carbon monoxide, using an instrument calibrated with gas mixtures containing known concentrations of carbon monoxide. The method as published in the appended "Federal Register" article was tested by means of a collaborative test involving a total of 16 laboratories. The test involved the analysis of both dry and humidified mixtures of carbon monoxide and air over the concentration range from 0 to 60 mg/cu m. A statistical analyais of the data of 15 laboratories is presented.

EVALUATION OF AN INTERLABORATORY COMPARISON INVOLVING PYROCARBON AND SILICON CARBIDE-COATED URANIUM-THORIUM CARBIDE BEADS

Usaec New Brunswick Laboratory, New Jersey

AUTHOR: Bingham, C. D.; Whichard, J.

ABSTRACT: An interlaboratory comparison program was conducted between six chemistry laboratories and three nondestructive assay laboratories. The material of interest was pyrocarbon- and silicon carbide-coated uranium-thorium carbide beads. Accuracy of uranium and thorium measurements was ascertained by supplying to the laboratories uranium oxide and thorium oxide samples containing known quantities. With one exception, the accuracy of the chemical analysis of uranium was within a range of 0.5% relative to the prepared value. Within-laboratory precisions ranged from 0.013 to 0.39% RSD for the mixed oxide samples. Chemical assay of the beads exhibited a range of nearly ±1% (relative) about the interlaboratory chemical average for uranium content. Within-laboratory precisions ranged from 0.03 to 0.38% RSD. Some dependence on sample preparation was evidenced. NDA measurements on mixed oxides showed biases as high as 3% from the prepared values. Measurements on coated beads were nearly comparable with chemical measurements in accuracy.

INTERLABORATORY EVALUATION OF THE TUNNEL TEST (ASTM E 84) APPLIED TO FLOOR COVERINGS

National Bureau of Standards, Washington, D.C. Institute for Applied Technology

AUTHOR: Lee, T. G.; Huggett, Clayton

ABSTRACT: Results of an interlaboratory evaluation of the ASTM E 84 tunnel test method involving eleven laboratories and nine materials, including four carpets, are reported. Data on flame spread, smoke, and fuel contribution are analyzed statistically. Selected physical characteristics of each tunnel are tabulated and compared relative to specifications in the test method. The between-laboratory coefficient of variation (reproducibility) in flame spread classification (FSC) was found to range from 7 to 29% for the four carpets and from 18 to 43% for the other materials tested. The between-laboratory coefficients of variation for smoke developed and fuel contribution ranged from 34 to 85% and from 22 to 117% respectively for all materials tested.

THE INTERLABORATORY PRECISION TEST. AN EIGHT LABORATORY EVALUATION OF THE PROVISIONAL ALGAL ASSAY PROCEDURE BOTTLE TEST

North Carolina University, Chapel Hill Department of Environmental Sciences and Engineering

AUTHOR: Weiss, Charles M.; Helms, Ronald W.

ABSTRACT: In order to establish the validity of an algal assay procedure for the determination of algal nutrient levels in surface waters, a suitable protocol was designed and followed by eight laboratories. This group consisted of one government laboratory, four university laboratories and three industrial laboratories. The basic procedure was to evaluate by use of the "bottle" or batch test the precision and reproducibility of the growth response of one test organism, Selenastrum capricornutum, in four media of varying nutrient strength. The medium was originally defined for the PAAP test and modified slightly in subsequent evaluations. The test media of this experiment were all dilutions of the PAAP medium.

INTERLABORATORY COMPARISON OF CHEMICAL ANALYSIS OF URANIUM MONONITRIDE

National Aeronautics and Space Administration, Lewis Research Center, Cleveland, Ohio

AUTHOR: Merkle, E. J.; Davis, W. F.; Halloran, J. T.; Graab, J. W.

ABSTRACT: Analytical methods were established in which the critical variables were controlled, with the result that acceptable interlaboratory agreement was demonstrated for the chemical analysis of uranium mononitride. This was accomplished by using equipment readily available to laboratories performing metallurgical analyses. Agreement among three laboratories was shown to be very good for uranium and nitrogen. Interlaboratory precision of ±0.04 percent was achieved for both of these elements. Oxygen was determined to ±15 parts per million (ppm) at the 170-ppm level. The carbon determination gave an interlaboratory precision of ±46 ppm at the 320-ppm level.

COOPERATIVE STUDIES ON LABORATORY METHODOLOGY

Journal American Water Works Association 51:607 (May 1959)

AUTHOR: Kramer, H. P.; Kroner, R. C.

ABSTRACT: The Analytical Reference of the Robert A. Taft Sanitary Engineering Center is a voluntary association of member organizations whose purpose is evaluation of methods in sanitary engineering. Samples are prepared to guarantee, to the extent possible, the desired concentrations of constituents. One aliquot was chosen at random and analyzed in the Sanitary Engineering Center to assure that no significant errors were made in sample preparation and to uncover possible difficulties not anticipated during sample design and preparation.

Sample Type I-A was the second sample for testing water mineral in approximately two years. The article summarized the results of studies made on sample Type I-A for calcium, magnesium, hardness, sulfate and chloride, alkalinity, sodium and potassium. Results obtained indicate that, in contrast to the determination of alkalinity, those of calcium, magnesium, hardness, sulfate, chloride, sodium, and potassium can be performed with a high degree of accuracy. The superiority shown by EDTA methods for hardness, calcium, and magnesium; and of the mecuric nitrate method for chloride was noted.

INTERLABORATORY COMPARISONS IN WATER BACTERIOLOGY

Journal American Water Works Association 59:237 (February 1967)

AUTHOR: Greenberg, A. E.; Thomas, J. S.; Lee, T. W.; Gaffey, W. R.

ABSTRACT: The article describes an experiment designed to examine whether test results differ between laboratories. The model used was a four-way, partially nested, mixed model analysis of variance in which laboratories, media, and water samples were assumed to be fixed effects, and days represented a random sample of days. The analysis of variance was performed for three separate interlaboratory comparisons.

This model makes it possible to evaluate main effects and interactions from one analysis. Test conclusions showed that results in the laboratory in question were acceptable. The results also showed several interactions which would bear followup.

CHEMICAL REFERENCE SAMPLES IN WATER LABORATORIES

Journal American Water Works Association 61:599 (November 1969)

AUTHORS: Greenberg, A. E.; Moskowitz, N.; Tamplin, B. R.; Thomas, J.

ABSTRACT: The article reports results of a single analysis of two water samples containing different ionic concentrations that were analyzed for the same constituents. Analytical results on both samples were received from 92 laboratories approved for chemical work. Youden's procedure for graphical diagnosis of interlaboratory test results was used, with some modifications, to evaluate the results from each laboratory. Circles defining acceptable, questionable, and unacceptable results were drawn. Thirty-eight laboratories had perfect scores, 54 had one or more unacceptable results; and two of the 54 had no acceptable results.

In 1961, 29 of 63 laboratories reported unacceptable results for one or more constituents. Results of the current test indicate there has been no general improvement in the intervening years. Lack of improvement was associated with an inadequate follow-up program.

USE OF REFERENCE SAMPLES IN EVALUATING WATER LABORATORIES

Public Health Reports 76:783 (September 1961)

AUTHOR: Greenberg, A. E.

ABSTRACT: A reference sample was used to evaluate sample results of approved water laboratories. Laboratories were sent replicate 1-gallon samples of water bottled at a water treatment plant handling surface water. Analysis for calcium, magnesium, sodium, potassium alkalinity, chloride, and sulphate were requested.

Analyses were to be made in duplicate. In the sanitation and radiation laboratory of the state health department, each of seven chemists analyzed the reference sample to provide basic information on its composition and the variability of results. Comparison of the approved laboratory results with those of the state health department laboratory showed four sources of variation in approved laboratories: (a) differences between replicate samples, (b) differences between laboratories, (c) differences between analysts, (d) differences between methods. A comparison of individual approved laboratories with all approved laboratories was made using results falling between the mean and ± 1 standard deviation as acceptable, results between ± 1 and ± 2 standard deviations from the mean were acceptable but questionable, and results outside the limits of ± 2 standard deviations from the mean were unacceptable. Twenty-nine of 63 participating laboratories, more than two-thirds, produced unacceptable results for one or more constituents. In summary, performance of a small number of laboratories was generally unacceptable. Performance of a larger number of laboratories was better, but occasionally unacceptable. With this information, the state health department laboratory instituted a follow-up program to rectify those laboratories needing improvement.

GRAPHICAL DIAGNOSIS OF INTERLABORATORY TEST RESULTS

Industrial Quality Control 15:24 (May 1959)

AUTHOR: Youden, W. J.

ABSTRACT: The article describes a double sample graphical analysis scheme for diagnosis of errors in interlaboratory test results. Samples of two different materials are sent to a number of laboratories which are asked to make one test on each material. The two materials should be similar and be reasonably close in the magnitude to the property evaluated. Diagnosis of the configuration of points makes possible identification of situations where more careful description or modification is required, erratic work, deviations from specified procedure, and prevalence of constant errors. A method for estimating standard deviation from test results is described. The graphical procedure facilitates presentation of the results in a convincing manner, thus avoiding statistical computations.

B. ANALYTICAL METHODS EVALUATION

COOPERATIVE EVALUATION OF TECHNIQUES FOR MEASURING HYDROCARBONS IN DIESEL EXHAUST

Cooridnating Research Council, Inc., New York

ABSTRACT: A small diesel engine was shipped to 13 laboratories in succession, and each laboratory measured exhaust hydrocarbon concentrations by methods of their own choosing. The standard deviation of the measured concentrations was on the order of 50% of the median values. Sources of the variation could be true differences in the exhaust samples from the engine, differences among laboratories in taking and handling the samples, and differences in instrument responses. Differences in sampling among laboratories appeared to be a major source of the variation.

COOPERATIVE EVALUATION OF TECHNIQUES FOR MEASURING HYDROCARBONS IN DIESEL EXHAUST, PHASE III

Coordinating Research Council, Inc., New York

ABSTRACT: Earlier cooperative tests indicated that errors in measuring hydrocarbon concentrations in diesel exhaust were undesirably large. To determine sources of the errors and to eliminate them, additional tests were conducted on one engine at a central location with twelve continuous hydrocarbon analyzers. Results of these tests show that with improvements in equipment and operating techniques, the precision and reliability of hydrocarbon measurements are satisfactory for current needs.

COORELATION OF FULL-FLOW LIGHT EXTINCTION TYPE DIESEL SMOKEMETERS BY A SERIES OF NEUTRAL DENSITY FILTERS

Coordinating Research Council, Inc., New York

ABSTRACT: The project involved testing twenty-four smokemeters by fourteen laboratories. The same series of four precalibrated metallic type neutral density filters was used by each laboratory in performing the static calibration of their diesel smoke measuring systems. The overall result was that essentially the same detector response was reported by the laboratories although each was asked to perform the static calibrations as they normally would. It may be concluded that the smokemeter results under static conditions are essentially equivalent and that no gross or consistent discrepancies could be found.

INSTRUMENTAL ANALYSIS OF CHEMICAL POLLUTANTS. TRAINING MANUAL

Environmental Protection Agency, Washington, D.C. Office of Water Programs

ABSTRACT: The manual was developed for use by students in training courses of the Water Quality Office, Environmental Protection Agency. The report discusses gas, liquid, and thin-layer chromatography, atomic and colorimetric spectral analysis, sampling methods, and instrument design. A special section for pesticide analysis of soil or water is also included.

METHODS FOR ORGANIC PESTICIDES IN WATER AND WASTEWATER

National Environmental Research Center, Cincinnati, Ohio.
Analytical Quality Control Laboratory

ABSTRACT: The report presents a general discussion, helpful hints and suggestions, and precautionary measures required for pesticide analysis. Step by step procedures are given for organochlorine pesticides.

EVALUATION OF MONITORING METHODS AND INSTRUMENTATION FOR HYDROCARBONS AND CARBON MONOXIDE IN STATIONARY SOURCE EMISSIONS

Walden Research Corporation, Cambridge, Massachusetts

ABSTRACT: The report reviews the state of the art of monitoring methods and instruments for carbon monoxide (CO) and hydrocarbons (HC) in stationary sources. Emissions are characterized from boilers, municipal incinerators, gray iron foundries, refineries, and asphalt batching plants. Manual methods for CO and HC determination are discussed, and monitoring instrumentation is reviewed. Nondispersive infrared spectroscopy (NDIR), gas chromatography, and flame ionization detection are evaluated in laboratory and pilot plant studies. Field evaluations were conducted on the reported industries. Calibration procedures, accuracy, and some results are reported. A computer program for data reduction is included.

C. STATISTICAL ANALYSIS - QUALITY CONTROL

COLLABORATIVE STUDY OF REFERENCE METHOD FOR DETERMINATION OF SULFUR DIOXIDE IN THE ATMOSPHERE (PARAROSANILINE METHOD)

Southwest Research Institute, Houston, Texas

AUTHOR: McKee, Herbert C.; Childers, Ralph E.; Saenz, Oscar Jr.

ABSTRACT: The report presents information obtained in the evaluation and collaborative testing of a reference method for measuring the sulfur dioxide content of the atmosphere. The technique is called the pararosaniline dye method or sometimes the West-Gaeke method. Different variations of this method have been used extensively by many laboratories since the original publication in 1956, and it has been found to be reliable and reasonably free of interferences. Collaborative tests were performed involving a total of eighteen laboratories. A statistical analysis of the data of fourteen laboratories provided the following results, based on the analysis of pure synthetic atmospheres using the 30-min sampling procedure and the sulfite calibration method prescribed. Results are also presented with respect to the use of control samples and reagent blank samples, the minimum number of samples required to establish validity of results within stated limits, and the statistical evaluation of various steps included in the method. The method can give satisfactory results only when followed rigorously by experienced laboratory personnel. The publication of the method in the Federal Register, April 30, 1971, as the reference method to be used in connection with Federal ambient air quality standards for sulfur dioxide is appended.

GUIDELINES FOR DEVELOPMENT OF A QUALITY ASSURANCE PROGRAM. REFERENCE METHOD FOR MEASUREMENT OF PHOTOCHEMICAL OXIDENTS

Research Triangle Institute, Durham, North Carolina

AUTHOR: Smith, Franklin; Nelson, A. Carl Jr.

ABSTRACT: Guidelines for the quality control of Federal reference method for photochemical oxidants are presented. These include: (1) good operating practices; (2) directions on how to assess data and qualify data; (3) directions on how to identify trouble and improve data quality; (4) directions to permit design of auditing activities; and, (5) procedures which can be used to select action options and relate them to costs. The document is designed for use by operating personnel.

GUIDELINES FOR DEVELOPMENT OF A QUALITY ASSURANCE PROGRAM. REFERENCE METHOD FOR THE CONTINUOUS MEASUREMENT OF CARBON MONOXIDE IN THE ATMOSPHERE

Research Triangle Institute, Research Triangle Park, North Carolina

ABSTRACT: The report has been prepared for the quality control of ambient air measurements of carbon monoxide. The purpose of the document is to provide uniform guidance to all EPA monitoring activities in the collection, analysis, interpretation, presentation, and validation of quantitative data. The technique used is non-dispersive infrared (NDIR) spectrometry.

COMPUTER-COUPLED QUALITY CONTROL PROCEDURE FOR GAMMA-RAY SCINTILLATION SPECTROMETRY

Naval Radiological Defense Lab, San Francisco, California

AUTHOR: Covell, D. F.

ABSTRACT: Long-term stabilization of instrumental performance is necessary for gamma-ray scintillation spectrometry whether used for nuclear spectroscopy studies or for radionuclide identification and estimation. This requirement is especially important if high-precision measurements are to be made on a routine basis. It is proposed to achieve sufficient stabilization through statistical quality control, a technique used to maintain the quality of output of a process or system. A quality control procedure was devised which consists of periodic measurement of a current standard spectrum and comparison of it, on a channel-by-channel basis on a computer, with a reference standard spectrum. Significant differences between the two spectra are interpreted as machine deviations that require correction. As part of the procedure, values obtained from this measurement are charted so that current and past performance can be compared easily. This makes possible a prompt awareness of unusual changes in performance. Application of the technique has resulted in improved stability, improved reliability, and reduced maintenance. Approximately 20 minutes of technician time are required per day to apply this procedure to a single instrument. Less time per instrument is required when several instruments are simultaneously controlled.

QUALITY CONTROL IN RADIOCHEMICAL ANALYSIS

Usaec Health and Safety Laboratory, New York; Woods Hole Oceanographic Institution, Massachusetts (Usa)

AUTHOR: Harley, J. H.; Volchok, H. L.

ABSTRACT: An ideal system of quality control in radiochemical analysis is described and some data relating to analysis of seawater are presented. Several basic factors which affect the quality of a radiochemical analysis are: the use of proper standards for calibration; the use of proper counter efficiencies and backgrounds; the proper determination of radiochemical recovery; correction of results for analytical blank, and the continual checking of the performance of the overall system for accuracy and precision.

HANDBOOK FOR ANALYTICAL QUALITY CONTROL IN WATER AND WASTEWATER LABORATORIES

National Environmental Research Center, Cincinnati, Ohio,

AUTHOR: Ballinger, D. G.; Booth, R. L.; Midgett, M. R; Kroner, R. C.; Kopp, J. F.

ABSTRACT: One of the fundamental responsibilities of management is the establishment of a continuing program to insure the reliability and validity of analytical laboratory and field data gathered in water treatment and wastewater pollution control activities. This handbook is addressed to laboratory directors, leaders of field investigations, and other personnel who bear responsibility for water and wastewater data. Subject matter of the handbook is concerned primarily with quality control for chemical and physical tests and measurements. Sufficient information is offered to allow the reader to inaugurate, or to reinforce, a program of analytical quality control which will emphasize early recognition, prevention and correction of factors leading to breakdowns in the validity of data.

LABORATORY QUALITY CONTROL MANUAL

Robert S. Kerr Water Research Center, Ada, Oklahoma

ABSTRACT: The Federal Water Pollution Control Administration (FWPCA) is concerned about laboratory quality and has initiated a program of improved effort in that direction. The manual deals with two areas of that program; statistical analytical quality control and record keeping. The manual describes statistical techniques as applied to analytical quality control. It is also concerned with record keeping as it applies to laboratory procedures and suggests a method of laboratory record keeping that should satisfy the most severe critic.

REVIEWS OF CURRENT LITERATURE ON ANALYTICAL METHODOLOGY AND QUALITY CONTROL, NUMBER 22

Battelle Columbus Laboratories, Ohio, Analytical Methodology Information Center

ABSTRACT: The report is a compilation of current literature in the field of water pollution methodology. The contents include physical and chemical methods, biological methods, microbiological methods, methods and performance evaluation, and instrument development.

GUIDELINES FOR DEVELOPMENT OF A QUALITY ASSURANCE PROGRAM. REFERENCE METHOD FOR THE DETERMINATION OF SULFUR DIOXIDE IN THE ATMOSPHERE

Research Triangle Institute, Durham, North Carolina

AUTHOR: Smith, Franklin; Nelson, A. Carl Jr.

ABSTRACT: Guidelines for quality control of the Federal reference method for sulfur dioxide are presented. These include: (1) good operating practices, (2) directions on how to assess and qualify data, (3) directions on how to identify trouble and improve data quality, (4) directions to permit design of auditing activities, (5) procedures for selecting action options and relating them to costs. This document is not a research report. It is for use by operating personnel.

FWPCA METHOD STUDY 1: MINERAL AND PHYSICAL ANALYSES

Federal Water Pollution Control Administration, Cincinnati, Ohio. Analytical Quality Control Laboratory

ABSTRACT: Pairs of synthetic water samples were prepared in three ranges of concentration for pH, specific conductance, total dissolved solids, total hardness, sodium, potassium, total acidity/alkalinity, chloride and sulfate for analysis by FWPCA Official Interim Methods for Chemical Analysis of Surface Waters. Fifty-one analysts from twenty laboratories in FWPCA and 5 non-FWPCA laboratories cooperated in this study. A statistical summary of the results indicates the precision and accuracy values obtainable in routine work.

WORKSHOP ON SAMPLE PREPARATION TECHNIQUES FOR ORGANIC POLLUTANT ANALYSIS HELD AT DENVER, COLORADO ON 2-4 OCTOBER 1973

National Field Investigations Center-Denver, Colorado

AUTHOR: Meiggs, Theodore O.

ABSTRACT: The emphasis of the workshop was placed upon the problems of sample collection, extraction, and fractionation prior to detection of the pollutants of interest by the appropriate detection techniques. Wherever possible, methods or procedures were stressed that were applicable to the analysis for general classes of organic compounds as opposed to procedures for individual compound identification. What follows is a summation of the techniques discussed at the workshop. Many of these are currently being used by water laboratories to analyze industrial effluents, natural waters, bottom sediments, and aquatic biota for industrial and agricultural organic-chemical pollutants. In addition, some discussion is provided regarding analytical quality control in the organic laboratory.

GUIDELINES FOR DEVELOPMENT OF A QUALITY ASSURANCE PROGRAM, VOLUME I. DETERMINATION OF STACK GAS VELOCITY AND VOLUMETRIC FLOW RATE (TYPE-S PITOT TUBE)

Research Triangle Institute, Durham, North Carolina

AUTHOR: Smith, Franklin; Wagoner, Denny E.; Nelson, A. Carl Jr.

ABSTRACT: The document presents guidelines for developing a quality assurance program for the determination of stack gas velocity and volumetric flow rate using a type-S pitot tube. The introduction lists the overall objectives for a quality assurance program and delineates the program components. The operations manual sets forth recommended operating procedures to assure the collection of data of high quality and instructions for performing quality control checks. The manual for a field team supervisor contains directions for assessing data quality on an intra-team basis and for collecting the information necessary to detect and/or identify trouble. The manual for manager of groups of field teams presents information relative to the test method (a functional analysis) to identify the important operations variables and factors, and statistical properties of and procedures for carrying out auditing procedures for an independent assessment of data quality.

STATISTICAL EVALUATION OF SELECTED ANALYTICAL PROCEDURES

Mound Laboratory, Miamisburg, Ohio

AUTHOR: Bohl, D. R.; Sellers, D. E.

ABSTRACT: A data evaluation study was conducted to evaluate the precision and accuracy of analytical procedures. Conventional statistical formulas were used to evaluate the data. The procedures evaluated statistically were a potentiometric method for determining iron and uranium, a volumetric titration of nickel, and the determination of uranium by controlled-potential colorimetric and potentiometric titration. The accuracy, standard deviation and confidence intervals were calculated using historical data from these procedures.

STANDARD REFERENCE MATERIALS: ANALYSIS OF INTERLABORATORY MEASUREMENTS ON THE VAPOR PRESSURES CF CADMIUM AND SILVER. (CERTIFICATION OF STANDARD REFERENCE MATERIALS 746 AND 748)

National Bureau of Standards, Washington, D.C. Institute for Materials Research (401 937)

AUTHOR: Paule, Robert C.; Mandel, John

ABSTRACT: Detailed statistical analyses have been made of results obtained from a series of interlaboratory measurements on the vapor pressures of cadmium and silver. Standard Reference Materials 746 (cadmium) and 748 (silver) which were used for the measurements have been certified over the respective pressure ranges 10 to the -11th to 10 to the -4th atm and 10 to the -12th to 10 to the -3rd atm. The temperature ranges corresponding to these pressures are 350-594 K for cadmium and 800-1600 K for silver. The heats of sublimation at 298 K and the associated two standard error limits for cadmium and silver are 26660 plus or minus 150 cal/mol and 68010 plus or minus 300 cal/mol, respectively. Estimates of uncertainty have been calculated for the certified temperature-pressure values as well as for the uncertainties expected from a typical single laboratory's measurements. The statistical analysis has also been made for both the second and third law methods, and for the within- and between-laboratory components of error. The uncertainty limits are observed as functions of both the heat of sublimation and the temperature.

HANDBOOK FOR ANALYTICAL QUALITY CONTROL IN WATER AND WASTEWATER LABORATORIES

National Environmental Research Center, Cincinnati, Ohio, Analytical Quality Control Laboratory

AUTHOR: Ballinger, D. G.; Booth, R. L.; Midgett, M. R.; Kroner, R. C.; Kopp, J. F.

ABSTRACT: One of the fundamental responsibilities of management is the establishment of a continuing program to insure the reliability and validity of analytical laboratory and field data gathered in water treatment and wastewater pollution control activities. This handbook is addressed to laboratory directors, leaders of field investigations, and other personnel who bear responsibility for water and wastewater data. Subject matter of the handbook is concerned primarily with quality control for chemical and physical tests and measurements. Sufficient information is offered to allow the reader to inaugurate, or to reinforce, a program of analytical quality control which will emphasize early recognition, prevention and correction of factors leading to breakdowns in the validity of data.

WATER METALS NO. 4, STUDY NUMBER 30. REPORT OF A STUDY CONDUCTED BY ANALYTICAL REFERENCE SERVICE

Public Health Service, Cincinnati, Ohio. Bureau of Prevention and Environmental Control

AUTHOR: McFarren, Earl F.; Parker, John H.; Lishka, Raymond J.

ABSTRACT: In the study, three samples containing between 0.005 and 5.0 mg per liter of each of nine metals - zinc, chromium, copper, magnesium, manganese, silver, lead, cadmium, and iron - were provided. Each participant was requested to do a single analysis for each of the metals in each of the three samples by the provided atomic absorption spectrophotometric method. This method, depending upon the sensitivity of the instrument (burner, tube, etc.) available, gave the participant a choice of aspirating the sample directly into the flame or of chelating with ammonium pyrrolidine dithiocarbamate and extracting into methyl isobutyl ketone before aspirating. The results obtained were evaluated in terms of whether the sensitivity of the method was sufficient to permit the measurement of the metal with a reasonable degree of precision and accuracy at the concentration prescribed by drinking water standards.

WATER SURFACTANT NO. 3, STUDY NUMBER 32. REPORT OF A STUDY CONDUCTED BY ANALYTICAL REFERENCE SERVICE

Public Health Service, Cincinnati, Ohio. Bureau of Disease Prevention and Environmental Control

AUTHOR: Lishka, Raymond J., Parker, John H.

ABSTRACT: In the study each participant was shipped three sterilized water samples in disposable 1-quart polyethylene containers. Sample 1 was composed of filtered river water containing 2.94 mg/liter linear alkylsulfonates (LAS). Sample 2 was tap water containing 0.48 mg/liter LAS. Sample 3 was distilled water containing 0.27 mg/liter LAS. A small amount of methylene blue and a copy of the procedure were sent with the samples. The data indicate no difference in methylene blue obtained from many different suppliers. Results from 111 analysts show good accuracy and precision for all samples.

THE QUALITY CONTROL OF LABORATORY PRECISION

American Journal of Clinical Pathology 25:585 (May 1955).

AUTHOR: Waid, M. E.; Hoffman, R. G.

ABSTRACT: The paper had four purposes; (1) to propose a method of using data of patients to evaluate the precision of laboratory procedures; (2) to illustrate the method with data from two general hospitals; (3) to fit frequency distribution curves to these data and illustrate their applicability; and (4) to demonstrate that the care of many patients may be affected by results that have been inaccurately standardized.

The best manner for using the method proposed in this paper is first to run standards of known concentration through the laboratory to insure that the laboratory is functioning properly. When assurance is gained that the laboratory is functioning properly, then the test results of the clinical specimens run during this same period may be used to set up the charts.

The steps in the method are: (1) The numerical value of each test is recorded. (2) All values for a particular test are added at the end of each day, or other predetermined period of time.
(3) Arithmetic means for each type of test are computed.
(4) The means obtained are plotted as points on a graph.
(5) Probability limits may be computed to be used as guidelines for the director of the laboratory.

Data on hemoglobin and red cell counts were tabulated for two general hospital laboratories. In one hospital, the hemoglobin tests were restandardized during the period covered by the data. In the other hospital, a suspected change in the hemoglobin level was seen. In both cases, the ability of the proposed method to portray these changes was graphically demonstrated.

The effects on medical practices which resulted from the hemoglobin restandardization were estimated by tabulating the number of patients who received transfusions. The transfusion rate was reduced approximately one-half.

Charts similar to those presented in this paper may be used for the control of any laboratory procedure.

ABSENCE OF ANALYTIC BIAS IN A QUALITY CONTROL PROGRAM

The American Journal of Clinical Pathology 38:468 (November 1962)

AUTHOR: Weinberg, M. S.; Barnett, R. N.

ABSTRACT: The article describes an experiment conducted to determine if the analyst produces incorrect results because of conscious or unconscious bias toward a known value, such as in a pool which may be used for several months, with all analysts aware of the anticipated results.

A single batch of pooled serums that had been in use and for which sufficient data on reliability had been accumulated was used in the study. Samples of the batch were used in routine daily quality control. Another sample was introduced as a blind sample during July and August in such a way as to prevent knowledge of such a sample by analysts.

An additional study was performed during the same period. Each technologist was instructed to choose one of the routine clinical samples for duplicate analysis for each determination requested.

In the study of blind versus known quality control serums introduced into routine clinical chemical determinations, no evidence was found that the analysts achieved a closer approach to the average known values nor a narrower 3 standard deviation range for the known samples. Values for duplicate determinations of unknown specimens were always closer than the comparative values of blind and known controls. The authors concluded that this was the result of more exact reproduction of analytic conditions rather than the effect of bias.

BIOMETRY: THE PRINCIPLES AND PRACTICE OF STATISTICS IN BIOLOGICAL RESEARCH

W. H. Freeman and Company, San Francisco

AUTHOR: Sokal, R. R.; Rohlf, F. J.

ABSTRACT: The abstract includes the table of contents and part of Appendix 3, Statistical Computer Programs. The computer programs included with a brief summary of their outputs is as follows:

A3.1 Basic statistics for ungrouped data. Output includes: mean, median, variance, standard deviation, coefficient of variation, g1, g2, and the Kolmogorov-Smirnov statistic Dmax resulting from a comparison of the observed sample with a normal distribution based on the sample mean and variance - followed by their standard errors and 100 (1 -) % confidence intervals where applicable.

A3.2 Basic statistics for data grouped into a frequency distribution. This program is similar to program A3.1, but is intended for data grouped into a frequency distribution.

A3.3 Goodness of fit to discrete frequency distributions. Options are provided for the following computations.

(1) Compute a binomial or poisson distribution with specified parameters.

(2) Compute the deviations of an observed frequency distribution from a binomial or poisson distribution of specified parameters or based on appropriate parameters estimated from observed data. A G-test for goodness of fit is carried out.

(3) A series of up to 10 observed frequency distributions may be read in and individually tested for goodness of fit to a specified distribution, followed by a test of homogeneity of the series of observed distributions.

(4) A specified expected frequency distribution (other than binomial or poisson) may be read in and used as the expected distribution. This may be entered in the form of relative expected frequencies or simply as ratios (for example 1:2:1). The maximum number of classes for all cases is thirty. In the binomial and poisson, the class marks cannot exceed $Y_i = 29$.

SYSTEMS CONTROL BY CUMULATIVE SUM METHOD

American Journal of Medical Technology 34:644 (November 1968)

AUTHOR: Griffen, D. F.

ABSTRACT: The article describes a system for plotting daily control data that is most useful where the secondary standards render recovery values on control or reference samples doubtful. The system involves subtracting an arbitrary target value from the daily recovery values of the control. Values for successive days are added algebraically to the previous day's total so a running difference from the target value is plotted. No actual confidence limit lines are drawn as parallels to the target or datum line. An out-of-control condition may be indicated by six successive climbing or falling plots, or when the cumulative sum track forms an angle of 45 degrees or greater on the datum line, if the linear distance between two successive vertical scale points is made equal to the linear distance between two successive horizontal points, and one such vertical scale segment is used as two standard deviations.

Trends and shift show up much more dramatically under this system of charting than they do on the usual X or $\bar{X}$ chart.

ESTABLISHING A QUALITY CONTROL PROGRAM FOR A STATE ENVIRONMENTAL LABORATORY

Water and Sewage Works, May 1974, pp 54 and ff

AUTHOR: Frazier, R. P.; Miller, J. A.; Murray, J. F.; Mauzy, M. P.; Schaeffer, D. J.; Westerhold, A. F.

ABSTRACT: The article describes five phases in development by the Illinois Environmental Protection Agency of a quality control program for regional laboratories. The current program is built around accuracy quality control charts. To develop these charts, every seventh sample entering the laboratory is divided into two portions, one of which is spiked by diluting with deionized water. By comparing new quality control data with that previously recorded, laboratory personnel are able to maintain a check on the analytical process.

At the end of each month, the quality control information consisting of the paired results on the original and spiked samples is assembled, entered on special data forms, and submitted to the data processing section for computer analysis. A summary report is distributed monthly to each of the laboratories. Using this information, the individual labs can take action for specific problems, while the division can take action for general problems.

The quality control program also uses externally prepared reference samples to provide an independent check of the various analyses.

Several support programs were intiated including checking the level of trace contaminants on bottles used for sample collection, field preservation, a standards group which prepares standards for the three laboratories, and an internal laboratory certification program to determine compliance of laboratories with procedures.

MANAGEMENT SYSTEM FOR AN ANALYTICAL CHEMICAL LABORATORY

American Laboratory S(1): 55 (January 1963)

AUTHOR: Krawczyk, D. F.; Byram, K. V.

ABSTRACT: The article describes a sample handling and verification system (SHAVES) to facilitate managing the analytical laborat[illegible]d to keep its records. It standardizes many laboratory procedures and automates many clerical tasks.

The principal elements of SHAVES are standardization, error checking, data reporting, and cost allocation. The system standardizes requests for analyses, recording of field data, reporting of laboratory analytical data, and limits of accuracy and precision for given determinations. The analyst must record all factors used in computing an analytical result, and the computer uses his factors to check it. The system detects errors in labeling and reporting of results. Costs for each analysis derived from the time and supplies required to perform it are available to the computer. The laboratory manager uses the monthly cost summary to make adjustments in financial support from programs using the laboratory. The system is near total effectiveness in detecting analytical computational errors.

THE NUMBER PLUS METHOD OF QUALITY CONTROL OF LABORATORY ACCURACY

The American Journal of Clinical Pathology 40:263 (September 1963)

AUTHOR: Hoffman, R. G.; Waid, M. E.

ABSTRACT: The number plus method uses clinical values as a source of quality control information. The procedure first involves obtaining a substantial number (about 500) of clinical values for the test in question, the organization of these values into a frequency distribution and then location of the mode. Next the percent of all tests that have values above the mode must be determined. Maintaining the order in which the 500 tests were made, they must be separated into groups of 50 consecutive tests. Then the number of tests that have values greater than the mode is counted. Plot the number of values that are greater than the mode (number plus) on a control chart. Control limits of any desired width can be constructed.

If the testing procedure was stable during the period over which the tests were made, then each group of 50 tests should have a number of plus tests (test values exceeding the mode) which be within the control limits. Control charts can be kept current by counting the number of plus test values as each group of 50 tests is completed. A point outside of the control limits, or a shift of values toward a control limit may indicate a shift in quality or control of clinical test results, and should be investigated.

Experience indicates that the procedure is sensitive enough to detect a shift of sufficient magnitude that it is worth looking for, but the shifts are small enough that they will not bias greatly the clinical use of given test results.

One advantage that the number plus method has over the reference standard method is: Number plus method uses clinical values, while a control serum frequently is not handled in the same manner as patient serums. Factors may influence a change in control serums or standards which do not apply to patient's specimens.

STANDARD DEVIATION: A PRACTICAL MEANS FOR THE MEASUREMENT AND CONTROL OF THE PRECISION OF CLINICAL LABORATORY DETERMINATIONS

The American Journal of Clinical Pathology 27:55/ (May 1957)

AUTHOR: Copeland, B. E.

ABSTRACT: Precision is defined as the closeness with which repeated analyses agree. The article describes the criteria for determining a measure of precision to include the following: The measure of precision can be used in common by all individuals interested in clinical precision - the pathologist, the technician, the clinician, the research scientist and the statistician. The data necessary for the calculation of the measure must be easy to collect, and the calculation must be easy to perform. The desired expression of precision must be easily interpreted. An exchange of letters or a personal interview should not be required to compare precision of one laboratory with the precision of another.

The standard deviation is described as a unit of precision which best fits the above critieria. A method for computing the standard deviation is described which uses the difference between duplicate measurements rather than differences from the mean.

Some of the conditions which must be stated to define adequately the frame of reference of the standard deviation are (1) number of technicians, (2) one or more days, (3) one or more samples, (4) concentration level of samples, (5) whether the technician knows he is being tested, etc.

EVALUATION OF A SYSTEM FOR PRECISION CONTROL IN THE CLINICAL LABORATORY

The American Journal of Clinical Pathology 48:243

AUTHOR: Barnett, R. N.; Pinto, C. L.

ABSTRACT: The article describes a method for quality control of clinical chemistry based on mixtures of patient samples. From each group of specimens submitted for analysis, two samples are selected and labeled A and B. Equal portions of A and B are mixed to form C, whose true value is (A + B)/2. Mixture C then becomes a sample which is analyzed with the other members of the batch. After all analyses are complete, the difference between the actual value for C and its theoretical value, (A + B)/2 is recorded. Forty such mixtures are analyzed on separate days and all the differences recorded. A control chart for these differences can then be prepared based on average deviations.

There are two disadvantages of the system in comparison with the conventional pooled plasma. It provides no check on accuracy. A change in reagents, standard, or procedure resulting in a shift of values is not detected. Because patient samples may exhibit values greatly different from those of a pool, the standard deviation determined from mixed samples might differ greatly from that of a pool merely because a different range of values was under study.

The above method was compared with the results found using frozen pooled serum for 10 commonly performed clinical chemical analyses. The standard deviations, coefficients of variation, and confidence limits were found to be close to those achieved by pooled serum technic. This substantiates the validity of limits of precision obtained by the use of serum pools.

SENSITIVITY - A CRITERION FOR THE COMPARISON OF METHODS OF TEST

Journal of Research, National Bureau of Standards 53:155 (February 1954)

AUTHOR: Mandell, J.; Stiehler, R. D.

ABSTRACT: In the evaluation of many methods of test, the two usual criteria - precision and accuracy - are insufficient. Accuracy is applicable only where comparisons with a standard can be made. Precision, when interpreted as degree of reproducibility, is not necessarily a measure of merit, because a method may be highly reproducible merely because it is too crude to detect small variations.

To obtain a quantitative measure of merit of test methods, a new concept, sensitivity, is introduced. If M is a measure of some property Q, and σ_M its standard deviation, the sensitivity of M, denoted by Ψ_M, is defined by the relation $\Psi_M = (dM/dQ)/\sigma_M$. It follows from this definition that the sensitivity of a test method may or may not be constant for all values of the property Q.

A statistical test of significance is derived for the ratio of the sensitivities of alternative methods of test. Unlike the standard deviation and the coefficient of variation, sensitivity is a measure of merit that is invariant with respect to any transformation of the measurement, and is therefore independent of the scale in which the measurement is expressed.

THE USE OF CONTROL CHARTS IN THE CLINICAL LABORATORY

American Journal of Clinical Pathology 20:1059 (1950)

AUTHOR: Levey, S.; Jennings, E.R.

ABSTRACT: The article describes a study of the use of control chart methods in a clinical laboratory. The control charts used were arithmetic mean $(\overline{X})$ and range (R). The method used whole blood and plasma in which the concentration of the substance estimated was stable over a long period, and in the range of normal blood values. Two samples each of whole blood and plasma were tested in the analysis twice a week. The true value of the concentration of any of the control substances was estimated by averaging the individual values obtained from the first 20 pairs analyzed over a period of about a month.

After the analysis was completed, the average and the range were plotted with the test value as ordinate and the order of test as abscissa. The statistical limits (three standard deviations) also were put on the chart.

Control charts were illustrated for urea nitrogen, plasma chloride, total plasma protein, plasma albumin, and carbon dioxide combining-power of plasma.

The control chart offers a simple method of checking the resultant effect of all factors influencing the accuracy of a test; e.g., the reagents, standards, time factors, technicians, and instruments used in the analysis. It offers a basis for action in initiating correction of a method that is not functioning properly. Also, it improves the general accuracy of a laboratory, because the technicians become control conscious and readily detect and report a test that is out of control. If the method is out of control, the chart usually cannot give the reason, and it is up to the analyst to determine the cause of the difficulty. Sometimes it is possible to note deterioration of reagents or standards by observing a trend in a control chart.

D. COMPUTER PROGRAMMING - INFORMATION RETRIEVAL

ASP - ANALYSIS OF SYNTHETICS PROGRAM FOR QUALITY CONTROL DATA

Du Pont de Nemours (E. I.) and Company, Aiken, South Carolina
Savannah River Laboratory

AUTHOR: Bailey, L. V.; Arnett, L. M.

ABSTRACT: The computer program, ASP, which calculates bias, precision, and other statistics of analytical methods, was written in FORTRAN IV for use on the IBM system/360-65. The Savannah River Plant laboratories use ASP monthly and quarterly to evaluate and to report the bias and precision of analyses important to process control and accountability.

A CHEMICAL INFORMATION CENTER EXPERIMENTAL STATION

Pittsburgh Chemical Information Center, Pennsylvania

AUTHOR: Arnett, E. M.

ABSTRACT: Reports are presented by the Principal Investigator and representatives of the following project task groups: library; programming; knowledge availability systems center; and behavioral research group. Each report is self-contained with its own abstract and appendices.

PROCEEDINGS. JOINT CONFERENCE ON PREVENTION AND CONTROL OF OIL SPILLS

American Petroleum Institute, New York

ABSTRACT: On December 15-17, 1969, a Joint Conference on Prevention and Control of Oil Spills was held under the co-sponsorship of the American Petroleum Institute and the Federal Water Pollution Control Administration. The objectives of the conference were to delineate the overall dimensions of the oil spills problem, explore the present state of the art of prevention and control of oil spills, and review the relevant research and development efforts of government and private industry, both here and abroad. The topics discussed include spill prevention, boom design, mechanical removal, chemical additives, analysis and sampling, monitoring, beach cleanup, fate of spills, ecological effects, and oil-spill information retrieval and dissemination.

OPERATIONAL HYDROMET DATA MANAGEMENT SYSTEM. DESIGN CHARACTERISTICS

North American Rockwell Information Systems Company, Anaheim, California

ABSTRACT: The hydromet system under development will include a central data bank operated by the U.S. Corps of Engineers, a large number of automated hydromet data gathering stations interfacing with the central data bank, and data retrieval facilities for interfacing the participating agencies with the data bank. The Operational Hydromet Data Management System (OHDMS) will be based in a large scale digital computer with appropriate large volume digital storage devices and peripherals. It will include a real-time digital data acquisition subsystem operating in association with an extensive manual data gathering network and a diverse user terminal subsystem for retrieval of stored hydromet data. The present study is structured to include the definition of the hardware and software characteristics of an integrated data management system to meet the requirements of each of the participating federal agencies. A key element in the study is the detailed definition of the user requirements for each of the federal participants.

DESIGN AND OPERATION OF AN INFORMATION CENTER ON ANALYTICAL METHODOLOGY

Battelle Memorial Institute, Columbus, Ohio. Columbus Laboratories

ABSTRACT: The report discusses the design and operation of a pilot analytical methodology information storage and retrieval system tailored to the needs of the Analytical Quality Control Laboratory (AQCL) and other segments of the National Analytical Methods Development Research Program (NAMDRP). All aspects of the system are presented.

STORAGE AND RETRIEVAL OF WATER QUALITY DATA. TRAINING MANUAL

Environmental Protection Agency, Washington, D.C. Water Quality

ABSTRACT: STORET is the data storage and retrieval system developed by and for the EPA and is a system suitable to the needs of all users of water quality and water resource data. The contents of the report make up a course which is intended to provide information and instruction on the STORET system for those persons directly involved in accumulating, processing and using water data.

STUDIES IN THE ANALYSIS OF METROPOLITAN WATER RESOURCE SYSTEMS. VOLUME V: A METHOD OF DATA REDUCTION FOR WATER RESOURCES INFORMATION STORAGE AND RETRIEVAL

Cornel University, Ithaca, New York. Water Resources and Marine Sciences Center

AUTHOR: Lewinger, K. L.

ABSTRACT: Data storage and retrieval expenses represent a significant portion of the cost of operating a management information system. The study focuses on the question of how much data already collected need be stored for future use, and on methods of reducing the quantity of data without necessarily reducing the information content. Several linear interpolation and least squares methods are explored for achieving data reduction, using as a means of illustration twenty-three different types of hydrologic records. Discussed also is the value of the data, the desired accuracy needed for various water resources studies, and the costs of data reduction as compared to data storage and retrieval.

PROCEEDINGS OF CONFERENCE ON "TOWARD A STATEWIDE GROUND WATER QUALITY INFORMATION SYSTEM" AND REPORT OF GROUND WATER QUALITY SUBCOMMITTEE, CITIZENS ADVISORY COMMITTEE, GOVERNORS ENVIRONMENTAL QUALITY CONTROL

Minnesota University, Minneapolis. Water Resources Research Center

ABSTRACT: The following topics were discussed: the natural quality of ground water in Minnesota, the use of ground water in Minnesota, hydrogeologic framework for deterioration in ground water quality, spray disposal of sewage effluent, solid waste disposal, needs and uses for a ground water quality data system, water well records and information system needs, subsurface geologic information system in Minnesota, ground water quality information system experiences in other states, Federal water information systems, and relation of ground water qualtiy information system and other systems in Minnesota.

AN INFORMATION SYSTEM FOR THE MANAGEMENT OF LAKE ONTARIO

Cornell University, Ithaca, New York

AUTHOR: Reynolds, Huey Dale

ABSTRACT: The first part of this study is concerned with a general analysis of information needs for the Experimental Operations Office (for Lake Ontario management) considering the purposes and objectives of the office, the boundary of the office, and the problem areas to be managed by the office. The second part deals with the theory of information and information systems in general, to provide a theoretical background. The third part consists of an analytical framework for an information system, followed by case studies of two particular areas, namely an economic base study and water quality control.

TRANSPORT AND THE BIOLOGICAL EFFECTS OF MOLYBDENUM IN THE ENVIRONMENT

Colorado State University, Fort Collins

ABSTRACT: The report presents an investigation of the transport and biological effects of molybdenum in the environment. The topics covered include: geochemistry of molybdenum, molybdenum transport in a reservoir, molybdenum toxicity studies in animals, fate of trace metals in a coal-fired power plant, molybdenum removal in conventional water and wastewater treatment plants, accumulation of available molybdenum in agricultural soils, levels of molybdenum in milk, analytical facilities, effects of dietary molybdenum on the physiology of the white rat, skeletal biology of molybdenum, information processing system, methodological problems in economic analysis of externalities and mineral development, perception of alternatives and attribution of responsibility for a water pollution problem, and information storage and retrieval routines.

DATA ACQUISITION SYSTEMS IN WATER QUALITY MANAGEMENT

Colorado State University, Fort Collins

AUTHOR: Ward, Rofer C.

ABSTRACT: The role of routine water quality surveillance was investigated, including a delineation of the objectives of a state water quality program based upon the state and federal laws. Seven specific objectives are listed under the two general objectives of prevention and abatement: planning, research, aid programs, technical assistance, regulation, enforcement, and data collection, processing, and dissemination. Each objective was broken down into the general activities required for its accomplishment and the data needed for each activity were identified. A survey of systems for grab sampling, automatic monitoring, and remote sensing was performed, each data acquisition technique being analyzed for capabilities, reliability, and cost. A procedure was developed for designing a state water quality surveillance program responsive to objectives. Financial and manpower constraints were considered.

THE SYSLAB SYSTEM FOR DATA ANALYSIS OF HISTORICAL WATER-QUALITY RECORDS (BASIC PROGRAMS)

Geological Survey, Washington, D.C.

AUTHOR: Steel, Timothy Doak

ABSTRACT: The report documents the basic computer programs comprising the SYSLAB system for systematically analyzing historical water-quality records. The first computer program retrieves station records for sets of water-quality variables from the survey's surface-water quality files. The procedure for analyzing water-quality data commonly has the following sequence: (1) a summary of basic statistics for each water-quality variable for the period of record or for shorter time increments, (2) plots of values of selected data pairs scaled according to the range of the data and (3) regression relationships based upon the graphic analysis of the plots. The appropriate SYSLAB computer program is given for each step in the sequence. Derivation of regression relationships is particularly applicable for the major inorganic chemical constituents which frequently are highly correlated with specific conductance. The report includes a description of the card set-up format and data input requirements for each computer program.

AUTOMATIC ACQUISITION OF WATER QUALITY DATA. A BIBLIOGRAPHY WITH ABSTRACTS

National Technical Information Service, Springfield, Virginia

AUTHOR: Lehmann, Edward J.

ABSTRACT: The NTISearch bibliography contains 51 selected abstracts of research reports retrieved using the NTIS on-line search system--NTISearch. The abstracts include the techniques and equipment used to obtain continuous water quality data. General system management and planning studies are covered.

FACTORS AFFECTING INNOVATION IN WATER QUALITY MANAGEMENT: IMPLEMENTATION OF THE 1968 MICHIGAN CLEAN WATER BOND ISSUE

Michigan University, Ann Arbor. Department of Civil Engineering

AUTHOR: Yaffee, Steven L.; Bulkley, Jonathan W.

ABSTRACT: This report focuses upon factors affecting innovation in the implementation of the 1968 Michigan Clean Water Bond Issue. The Joint Legislative Committee on Water Resources Planning which sized the bond program did not consider nutrient removal or any treatment beyond secondary in its determination of the fiscal resources necessary to meet 1980 Water Pollution Control objectives. Consequently, the fiscal resources were limited from inception. The net effect of the Clean Water Bond program maintains a 1968 status quo situation. Factors resisting innovation are identified and factors enhancing innovation are identified. An automated information storage/retrieval system for monitoring wastewater treatment facility funding is developed. Structural and process changes for future innovation are recommended.

MICHIGAN WATER RESOURCES ENFORCEMENT AND INFORMATION SYSTEM

Michigan Department of Natural Resources, Lansing. Water Resources Commission

AUTHOR: Guenther, Gary; Mincavage, Daniel; Morley, Fred

ABSTRACT: The project demonstrated an interactive federal/state water pollution control, enforcement, and information system, including interactive computer graphics as a method of output presentation. Two systems were interfaced: Michigan's Water Information System for Enforcement (WISE) and EPA's STORET system The WISE system is used to alert enforcement personnel to problems through exception reporting, and to provide follow-up information on these problems. STORET is used as a storage and retrieval system for water quality and inventory information. As information enters WISE, certain inputs are coded for storage in STORET. The interface mechanism is a common numbering system. Because WISE is modular in design, it can be used in part or in total by other agencies. The demonstration indicated that careful consideration should be given to the information that will comprise the computer file. Administrative, procedural, and auditing techniques should be completely set down before proceeding with management's commitment to the system. Microfilm should be used when feasible, both as Computer Output Microfilm (COM) and in manual files.

A NATIONAL OVERVIEW OF EXISTING COASTAL WATER QUALITY MONITORING

Interstate Electronics Corporation, Anaheim, California Oceanics Division

ABSTRACT: An overview of coastal water quality monitoring activity is presented, including an examination of related factors such as water quality standards, population, waste discharges, ocean dumping, a survey of data banks at the national level and others. Data from several inventories pertinent to coastal zone water quality is summarized to the state and EPA regional level with extensive descriptions contained in appendices.

SIDES: STORET INPUT DATA EDITING SYSTEM

Environmental Protection Agency, Athens, Georgia Surveillance and Analysis Division

AUTHOR: Barrow, David R.

ABSTRACT: The Water Quality Control Information System provides a broad data management capability for all activities of EPA's water programs activities. Central to both the program activities and the data management system is the need to store and retrieve ambient water quality data. The initial stages of the data management system were designed to fulfill that basic need. That was the beginning of STORET. The present report provides documentation for SIDES, a procedure designed specifically for field survey data and medium speed terminal, card input applications.

THERMOPHYSICAL AND ELECTRONIC PROPERTIES INFORMATION ANALYSIS CENTER (TEPIAC): A CONTINUING SYSTEMATIC PROGRAM ON TABLES OF THERMOPHYSICAL AND ELECTRONIC PROPERTIES OF MATERIALS

Purdue University, Lafayette, Indiana Thermophysical Properties Research Center

AUTHOR: Ho, Cho-Yen

ABSTRACT: The final report describes the activities and accomplishments of the Thermophysical and Electronic Properties Information Analysis Center (TEPIAC), which comprises internally the Thermophysical Properties Information Analysis Center (TPIAC) and the Electronic Properties Information Center (EPIC). TEPIAC's activities reported herein include literature search, acquisition, review, and codification; substance classification and organization; operation of a computerized information storage and retrieval system; publication of the Thermophysical Properties Research Literature Retrieval Guide Supplement; data extraction and compilation; data evaluation, correlation, analysis, synthesis, and generation of recommended reference values; publication of the TPRC Data Series, state-of-the-art summaries, and critical reviews; technical and bibliographic inquiry services; and current awareness and promotion efforts. TPIAC covers 14 thermophysical properties of all matter at all temperatures. EPIC covers 22 electronic (including also electrical, magnetic, and optical) properties and property groups of selected material groups at all temperatures.

STORET II: STORAGE AND RETRIEVAL OF DATA FOR OPEN WATER AND LAND AREAS

Federal Water Pollution Control Administration, Washington, D.C. Division of Pollution Surveillance

AUTHOR: Dubois, Donald P.

ABSTRACT: STORET Subsystem II described in this manual consists of a series of related computer programs designed for the efficient storage and retrieval of data collected in connection with water quality management programs. The system is intended for use in handling data collected from large open bodies of water and from points on land areas which cannot be associated readily with points on a stream.

PART I. A CONCEPTUAL MODEL FOR A TERRESTRIAL ECOSYSTEM PERTURBED WITH SEWAGE EFFLUENT, WITH SPECIAL REFERENCE TO THE MICHIGAN STATE UNIVERSITY WATER QUALITY MANAGEMENT PROJECT. PART II. A PERSONALIZED BIBLIOGRAPHIC RETRIEVAL PACKAGE FOR RESOURCE SCIENTISTS

Michigan State University, East Lansing. Department of Fisheries and Wildlife

AUTHOR: Conley, Walt,; Tipton, Alan R.

ABSTRACT: The report is provided in two distinct but interconnected parts. Part I contains discussions of management and design problems, components of terrestrial ecosystems, and specific site descriptions, all as they pertain to the sewage effluent spray program of the Michigan State University Water Quality Management Project. Part II began as an effort to compile a bibliographic reference file for the above project. This portion grew into the construction of relevant software, and was built around a 2500 citation bibliography. The bibliography is specifically oriented towards sewage effluent treatments, and is currently operative and available for interested researchers. A second bibliography is also described in this section.

SECTION XII

APPENDIX

SAMPLE LETTER

Directors of EPA Environmental Laboratories

Gentlemen:

Currently the Environmental Monitoring and Support Laboratory (EMSL-Cincinnati) has issued a contract for the "Development of a System for Conducting Inter-Laboratory Tests for Water Quality and Effluent Measurements." A pilot test program is being conducted to evaluate the validity of the inter-laboratory test program proposed by the Contractor.

Within the next month (mid-November) you will receive six chemical reference samples which are being distributed to the 22 EPA laboratories engaged in environmental monitoring.

The constituents to be determined are aluminum, arsenic, cadmium, copper, iron, mercury, lead, maganese, nickel, selenium, zinc, and cobalt. Laboratories should analyze for all these constituents.

The attached table provides an approved list of the standard methods for the chemical analysis of water. It is assumed that atomic absorption spectroscopy will be used where it is available and appropriate for a given element. Since the concentration of metals in at least one of the samples may be below the limit of detection to determine these levels, some form of concentration procedure such as chelation and extraction with organic solvents must be employed before analysis if flameless atomization is not used.

The sample should be analyzed as received; no dilution is required. A reporting form is enclosed.

TABLE 12-1. STANDARD METHODS FOR CHEMICAL ANALYSIS OF WATER: LIST OF APPROVED TEST PROCEDURES*

Parameter (mg/1)	Method	References (page numbers)		
		Standard methods	ASTM	EPA methods
Analytical methods for trace metals:				
Aluminum-total	Atomic absorption	210		98
Aluminum-dissolved	0.45 micron filtration and reference method for total aluminum			86
Antimony-total	Atomic absorption			
Antimony-dissolved	0.45 micron filtration and reference method for total antimony			86
Arsenic-total	Digestion plus silver diethyldithiocarbamate; atomic absorption	65,62		13
Arsenic-dissolved	0.45 micron filtration and reference method for total arsenic			86
Barium-total	Atomic absorption	210		
Barium-dissolved	0.45 micron filtration and reference method for total barium			86
Beryllium-dissolved	Aluminon; atomic absorption	67,210		
Beryllium-dissolved	0.45 micron filtration and reference method for total beryllium			86
Boron-total	Curcumin	69		
Boron-dissolved	0.45 micron filtration and reference method for total boron.			86
Cadmium-total	Atomic absorption; colorimetric	210,422	692	101
Cadmium-dissolved	0.45 micron filtration and reference method for total cadmium			86
Calcium-total	EDTA titration; atomic absorption	84	692	102
Calcium-dissolved	0.45 micron filtration and reference method for total calcium			86
Chromium VI	Extraction and atomic absorption; colorimetric	429		94
Chromium VI-dissolved	0.45 micron filtration and reference method for total chromium VI			86
Chromium-total	Atomic absorption; coloimetric	210,426	692,403	104

---

* Federal Register, Vol. 40, No. 111, Monday, June 9, 1975.

TABLE 12-1 (CONTINUED). STANDARD METHODS FOR CHEMICAL ANALYSIS OF WATER: LIST OF APPROVED TEST PROCEDURES

Parameter (mg/l)	Method	References (page numbers) Standard methods	ASTM	EPA methods
Chromium-dissolved	0.45 micron filtration and reference method for total chromium			86
Cobalt-total	Atomic absorption		692	
Cobalt-dissolved	0.45 micron filtration and reference method for total cobalt			86
Copper-total	Atomic absorption; colorimetric	210,430	692,410	106
Copper-dissolved	0.45 micron filtration and reference method for total copper			86
Gold-total	Atomic absorption			
Iridium-total	-----do			
Iron-total	Atomic absorption; colorimetric	210,433	692,152	108
Iron-dissolved	0.45 micron filtration and reference method for total iron			86
Lead-total	Atomic absorption; colorimetric	210,436	692	110
Lead-dissolved	0.45 micron filtration and reference method for total lead			86
Magnesium-total	Atomic absorption; gravimetric	210,416,201	692	112
Magnesium-dissolved	0.45 micron filtration and reference method for total magnesium			86
Manganese-total	Atomic absorption	210	692	114
Manganese-dissolved	0.45 micron filtration and reference method for total manganese			86
Mercury-total	Flameless atomic absorption			
Mercury-dissolved	0.45 micron filtration and reference method for total mercury			86
Molybdenum-total	Atomic absorption			
Molybdenum-dissolved	0.45 micron filtration and reference method for total molybdenum			86
Nickel-total	Atomic absorption; colorimetric	443	692	
Nickel-dissolved	0.45 micron filtration and reference method for total nickel			86
Osmium-total	Atomic absorption			
Palladium-total	-----do			
Platinum-total	-----do			

TABLE 12-1 (CONTINUED). STANDARD METHODS FOR CHEMICAL ANALYSIS OF WATER: LIST OF APPROVED TEST PROCEDURES

TABLE 12-1 (Continued). STANDARD METHODS FOR CHEMICAL ANALYSIS OF WATER: LIST OF APPROVED METHODS

Parameter (mg/l)	Method	References (page numbers)		
		Standard methods	ASTM	EPA methods
Potassium-dissolved	Atomic absorption; colorimetric; flame photometric	283,285	326	115
Potassium-dissolved	0.45 micron filtration and reference method for total potassium			86
Rhodium-total	Atomic absorption			
Ruthenium-total	-----do			
Selenium-total	-----do			
Selenium-dissolved	0.45 micron filtration and reference method for total selenium			86
Silica-dissolved	0.45 micron filtration and molybdosilicate-colorimetric	303	83	86,273
Silver-total	Atomic absorption	210		
Silver-dissolved	0.45 micron filtration and reference method for total silver			86
Sodium-total	Flame photometric; atomic absorption	317	326	118
Sodium-dissolved	0.45 micron filtration and reference method for total sodium			86
Thallium-total	Atomic absorption			
Thallium-dissolved	0.45 micron filtration and reference method for total thallium			86
Tin-total	Atomic absorption			
Tin-dissolved	0.45 micron filtration and reference method for total tim			86
Titanium-total	Atomic absorption			
Titanium-dissolved	0.45 micron filtration and reference method for total titanium			86
Vanadium-total	Atomic absorption, colorimetric	357		
Vanadium-dissolved	0.45 micron filtration and reference method for total vanadium			86
Zinc-total	Atomic absorption; colorimetric	210,444	692	120
Zinc-dissolved	0.45 micron filtration and reference method for total zinc			86